Progress In Electromagnetics Research, Vol. 175, 45–79, 2022

Machine Learning-Assisted Sensing Techniques for Integrated Communications and Sensing in WLANs: Current Status and Future Directions

Siyuan Shao[1], Min Fan[1], Chen Yu[1], Yan Li[3, 4], Xiaodong Xu[3, 4], and Haiming Wang[1, 2, *]

Abstract—Sensing is a key basis for building an intelligent environment. Using channel state information (CSI) from the IEEE 802.11 physical layer in the wireless local access networks, the CSI-based device-free sensing technique has become very promising to the current sensing solutions because of its non-invasion of privacy, non-contact, easy deployment, and low cost. In recent years, the integrated communication and sensing (ICAS) technology has become one of the popular research topics in both wireless communications and computer areas. Given the fruitful advancements of ICAS, it is essential to review these advancements to synthesize and give previous research experiences and references to aid the development of relevant research fields and real-world applications. Motivated by this, this paper aims to provide a comprehensive survey of CSI-based sensing techniques. This study categorizes the surveyed works into model-based methods, data-based methods, and model-data hybrid-driven methods. Some important physical models and machine learning algorithms are also introduced. The sensing functions are classified into detection, estimation, and recognition according to specific application scenarios. Furthermore, future directions and challenges are discussed.

1. INTRODUCTION

Wireless mobile communication is a driving force for the Internet of Everything (IoE) [1]. The integrated communication and sensing (ICAS) has become a very promising candidate for future wireless local access network (WLAN). Wi-Fi 6, as the latest generation of commercial Wi-Fi standards, is the first one specially designed for the IoE compared to the previous five generation versions based on the IEEE 802.11 standards [2]. Nowadays, wireless communication has moved from basic human-computer interaction to intelligent IoE. Various emerging industries and technologies, such as augmented reality (AR), smart home, and holographic interaction, not only put forward new requirements for communication performance but also bring communication into the framework of the Internet of Things (IoTs) [3–5].

As the key to building an intelligent environment, sensing is the foundation of the intelligent connection of all things. Compared with the conventional IoT, which uses a variety of sensors, wearable external devices, or cameras in computer vision to achieve sensing and interaction functions, wireless communication system uses the ubiquitous electromagnetic waves in space as a medium to achieve communication functions and has developed various sensing functions [6, 7]. This ICAS technology has received wide attention in the past decade. Researches on various application scenarios emerge in an endless stream. Among them, the indoor sensing research work with wireless Access Points (APs) or Wi-Fi as communication and sensing devices is particularly effective [8, 9].

Received 29 April 2022, Accepted 27 July 2022, Scheduled 9 August 2022
* Corresponding author: Haiming Wang (hmwang@seu.edu.cn).
[1] State Key Laboratory of Millimeter Waves, Southeast University, Nanjing 210096, China. [2] Pervasive Communication Research Center, Purple Mountain Laboratories, Nanjing 211111, China. [3] China Mobile Research Institute, Beijing 100053, China. [4] China Mobile Communications Group Co., Ltd, Beijing 100053, China.

The CSI-based schemes do not rely on the Line-of-Sight (LoS) path and can realize anytime, anywhere sensing in a non-contact manner. At the same time, due to the widespread deployment of WLAN and the maturity of its technology, the cost of CSI-based schemes becomes much lower. Therefore, implementing the sensing function on Commercial-Off-The-Shelf (COTS) Wi-Fi devices, and then realizing the ICAS has great application value, which will bring unprecedented convenience to our lives.

In the physical layer of the WLAN, the CSI under Multiple-Input and Multiple-Output (MIMO) and Orthogonal Frequency Division Multiplexing (OFDM) technologies is the channel frequency response (CFR) on each subcarrier between each pair of transmitting and receiving antennas. It reflects the multipath, shadow, and diffraction experienced by each subcarrier signal from the transmitting antenna to the receiving antenna. That is, the influence of the environment on electromagnetic waves will be manifested in CSI. Therefore, through the analysis of CSI, environmental information can be extracted, and then the purpose of sensing can be achieved. However, this environmental information cannot be presented intuitively. One solution is to construct a special model and make the special transformations of the CSI so that it can reflect some dynamic changes in the environment. Another way is to introduce artificial intelligence (AI). A large amount of measured data provides high-quality data sources for AI. In ICAS, most sensing problems can be boiled down to detection, recognition, and estimation problems [9], which is what AI is good at. Numerous studies have shown that many perceptual problems that cannot be solved by traditional modeling methods in ICAS can be addressed using machine learning methods.

This paper will focus on the CSI-based ICAS in WLAN and introduce the commonality and general methods of ICAS research work so far. Meanwhile, some key issues and future development directions are present. In Section 1, we outline the concepts and applications of ICAS. In Section 2, we mainly introduce the basic concepts of CSI and the devices and standards that are currently widely used in Wi-Fi to implement ICAS. In Section 3, we summarize and categorize the methods used in most of the current literature and highlight some important general methods. In Section 4, we analyze the application of ICAS in combination with the literature. In Section 5, we present the main difficulties and bottlenecks in the current CSI-based ICAS and propose solutions and prospects for future work.

2. BACKGROUND OF WLAN ICAS

Wi-Fi is a WLAN technology based on the IEEE 802.11 standard. Connectable network devices are wirelessly connected through wireless APs. Low latency and high transmission have always been the premise of the IoE. The latest Wi-Fi 6 technology, based on the IEEE 802.11ax protocol, has nearly three times the actual rate compared to the previous generation of Wi-Fi 5, and the theoretical rate even reaches 9.6 Gbps [10]. The access volume is four times that of the previous generation. The application of Orthogonal Frequency Division Multiple Access (OFDMA) technology enables the network to withstand low-latency access of more terminals, and the Multi-User Multiple-Input Multiple-Output (MU-MIMO) technology enables wireless APs to provide greater system capacity. Target Wake Time (TWT), an improved timing wake-up mechanism from 802.11ah, provides Wi-Fi 6 with an important resource scheduling function. The terminal can be woken up by the AP at a specific time by establishing a "wake-up protocol" with the AP. This greatly saves the power consumption of terminal devices and makes IoT devices more intelligent [11]. Nowadays, Wi-Fi technology provides a very friendly platform for building smart environments. WLAN is also moving closer to the general trend of the future intelligent IoE [12].

We will briefly analyze and explain the CSI extracted from WLAN and summarize and introduce the main standards and devices used in the current CSI-based ICAS technology research.

2.1. RSSI and CSI

Wi-Fi-based systems analyze the changes in wireless signal characteristics, such as Received Signal Strength Index (RSSI) or CSI, caused by reflection and diffraction phenomena by people or objects. RSSI from the MAC layer is a measure of the power of the received radio signal. There is no strict definition of RSSI in the IEEE 802.11 standard. Its magnitude is mainly affected by path loss, occlusion, and multipath effects. It reflects the channel quality to a certain extent. However, since its measurement

is the result of the superposition of all multipaths, the spatial information becomes blurred after the signals undergoing different attenuation, reflection, and refraction are superimposed. At the same time, due to the influence of multipath, the RSSI is less stable, and its value will fluctuate greatly even in static scenes. Therefore, RSSI can only provide coarse-grained information about the communication link, which greatly limits its potential in terms of sensing.

CSI from the physical layer contains the CFR information of each subcarrier between each pair of transmitting and receiving antennas in the communication system. And the CSI between the n-th transmitting antenna and the m-th receiving antenna can be expressed in the following form,

$$\mathbf{H}_{m,n} = [h_{m,n,1}, h_{m,n,2}, \ldots, h_{m,n,S}]^{\mathrm{T}}, \tag{1}$$

where S is the total number of subcarriers, and $h_{m,n,s}$ with $s = 1, 2, \ldots, S$ is the complex CFR in the s-th subcarrier,

$$h_{m,n,s} = |h_{m,n,s}| \exp(j\angle h_{m,n,s}), \tag{2}$$

with $|h_{m,n,s}|$ and $\angle h_{m,n,s}$ being the amplitude and phase. The OFDM technology used in Wi-Fi allocates signals to each orthogonal subcarrier for transmission. Therefore, each subcarrier corresponds to one CFR in the CSI. The CSI combination of all transmitting and receiving antenna pairs can be expressed as an $MS \times N$ CSI matrix as follows,

$$\tilde{\mathbf{H}} = \begin{bmatrix} \mathbf{H}_{1,1} & \mathbf{H}_{1,2} & \ldots & \mathbf{H}_{1,N} \\ \mathbf{H}_{2,1} & \mathbf{H}_{2,2} & \ldots & \mathbf{H}_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{H}_{M,1} & \mathbf{H}_{M,2} & \ldots & \mathbf{H}_{M,N} \end{bmatrix}, \tag{3}$$

where M and N are the numbers of receiving and transmitting antennas, respectively.

In the practical environments, the CSI can be expressed in the form of multipaths superposition. The spatial experience of each subpath is imprinted in the CSI. Certain environmental information can be extracted by analyzing the collected CSI. Compared with RSSI, which only provides the signal power at the receiving end, the CSI matrix is extended in the subcarrier domain and uses phase and amplitude to describe more abundant environmental information. Therefore, CSI can provide fine-grained information of the spatial environment.

2.2. Standards and Devices

The acquisition of RSSI is relatively simple. Most chips take into account the collection of RSSI, so RSSI information can be easily obtained on Wi-Fi. And for CSI information, not all network interface controllers (NICs) provide CSI data. The CSI information must be obtained by modifying the driver and using some specific COTS device NICs or using software-defined radio (SDR).

As the standard of Wi-Fi 4, IEEE 802.11n works in two frequency bands, 2.4 GHz and 5 GHz. There are two options for bandwidth, 20 MHz and 40 MHz, and the number of subcarriers is 56 and 114 respectively. Up to 64QAM and 4×4 spatial streams are supported. Since Wi-Fi 4 was released in 2009, many 802.11n-based measurement and experimental platforms have been developed so far, which can be used to measure only need a specific NIC. The 802.11n CSI Tool was developed by Halperin et al. of the University of Washington in 2011 [13]. Running it in Linux system with Intel 5300 Wi-Fi NIC can obtain the CSI information of the 802.11n standard. But this tool can only provide compressed CSI data of 30 subcarriers for both 20 MHz and 40 MHz bandwidths. The Atheros CSI Tool developed by Yaxiong Xie et al. in 2015 is also an open-source 802.11n measurement tool [14]. Using this tool, detailed physical layer information, including CSI information, can be obtained from the Atheros Wi-Fi NIC. This tool can provide complete subcarrier CSI data including 56 sub-carriers at 20 MHz and 114 sub-carriers at 40 MHz. This method of modifying the driver to use special tools and NICs can easily obtain 802.11n CSI information, so most CSI-based ICAS research chooses to collect experimental data in this way.

Additionally, the SDR platform, such as USRP and WARP, can also be used to obtain the CSI information of the received signal. 802.11n CSI Tool and Atheros CSI Tool only support CSI acquisition in the 802.11n standard. Using SDR can measure CSI of higher bandwidth standards, such as 802.11ac used in Wi-Fi 5, which operates in the 5 GHz band and has higher 80 MHz and 160 MHz bandwidths.

The number of subcarriers corresponding to the two bandwidths reaches 242 and 484 respectively. The SDR platform provides more options for CSI-based research work. At present, there have been many ICAS works based on USRP [15–19] or WARP [20].

3. METHODS AND ALGORITHMS

Extract fine-grained information from the received CSI raw data, and then achieve the purpose of sensing the environment. Special processing or learning of CSI is the commonly used method. This section categorizes the surveyed literature into model-based methods, data-based methods, and model-data hybrid-driven methods, according to whether special physical models are built, or machine learning is used. And the commonly used models and algorithms are explained for reference in the research work of ICAS.

3.1. Model-Based Methods

It was mentioned in the previous section that CSI is a three-dimensional matrix. The three dimensions correspond to the transmitting antenna, the receiving antenna, and the subcarriers, respectively. The elements with superscript s and subscripts n, m can be expressed in the form of multipaths superposition as follows,

$$h_{m,n,s} = \sum_{l=1}^{L_{m,n}} \alpha_{m,n,s,l} e^{-j2\pi f_s \frac{d_{m,n,s,l}}{c}}, \tag{4}$$

where $L_{m,n}$ is the total number of multipaths from the n-th transmitting antenna to the m-th receiving antenna; l is the l-th subpath; c is the propagation velocity of the electromagnetic wave in space; α and d are the path loss and length of the l-th subpath of the s-th subcarrier from the n-th transmitting antenna to the m-th receiving antenna, respectively; f_s is the frequency of the s-th subcarrier.

It can be seen that the CFR expressed by Eq. (4) contains most of the information on multipath in space. These multipaths include the highest energy LoS path, constant static, and changing dynamic object reflection and diffraction paths. This information is very helpful for sensing the environment. But after these subpaths are superimposed on each other, they only appear as a CFR in the form of a pair of amplitude and phase. Multiple CFRs are combined into the form of CSI. At this time, it is necessary to establish some special models and extract the environmental information from them through some transformations.

3.1.1. Doppler Spread Model

Doppler spread model is a model commonly used to study the speed of moving objects in an environment. Doppler spread is the spectral spreading caused by the time rate of change of the radio channel. In communication, it usually refers to the situation in which the frequency of the received signal is shifted because the radio transmitting antenna and receiving antenna are in a state of relative motion. The faster the relative speed is, the more severe the Doppler spread is. In the ICAS system, the positions of the transmitting and receiving antennas are usually fixed. Doppler spread in this scenario refers to the situation where there is a moving object between the transmitting and receiving antennas that changes the length of the signal path, thereby causing the phase of the signal at the receiving antenna to shift, as shown in Fig. 1. Eq. (4) represents the CFR in a static scene. When there are moving objects in the environment that affect the signal transmission, Eq. (4) becomes the following form,

$$\begin{aligned} h_{m,n,s}(t) &= \sum_{l=1}^{L^{\mathrm{ST}}_{m,n}} \alpha_{m,n,s,l} e^{-j2\pi f_s \frac{d^{\mathrm{ST}}_{m,n,s,l}}{c}} + \sum_{l=1}^{L^{\mathrm{DY}}_{m,n}} \beta_{m,n,s,l}(t) e^{-j2\pi f_s \frac{d^{\mathrm{DY}}_{m,n,s,l}(t)}{c}} \\ &= h^{\mathrm{ST}}_{m,n,s} + \sum_{l=1}^{L^{\mathrm{DY}}_{m,n}} \beta_{m,n,s,l}(t) e^{-j2\pi f_s \frac{d^{\mathrm{DY}}_{m,n,s,l}(t)}{c}}, \end{aligned} \tag{5}$$

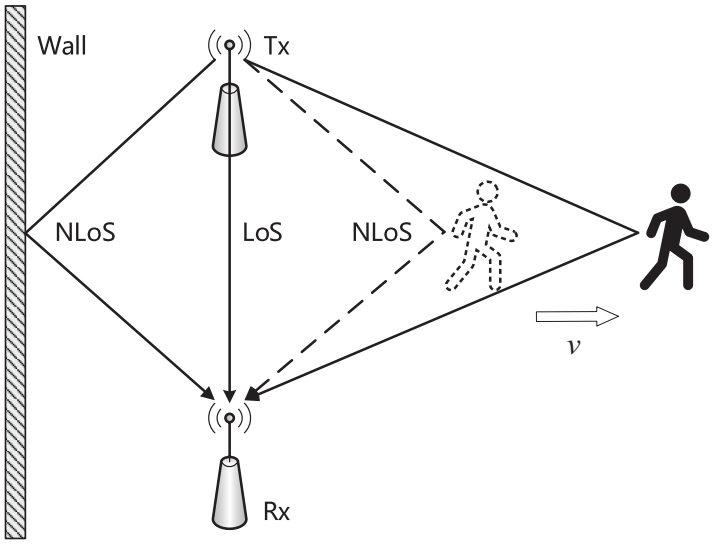

Figure 1. Phase shift of the received signal caused by the moving object changing the subpath length.

where $h^{\mathrm{ST}}_{m,n,s}$ is the superposition of the static subpaths; the rest is the superposition of the dynamic subpaths caused by the path loss and path length of some subpaths changed by the moving object; $L^{\mathrm{ST}}_{m,n}$ is the number of static subpaths; $L^{\mathrm{DY}}_{m,n}$ is the number of dynamic subpaths; $\beta_{m,n,s,l}(t)$ and $d^{\mathrm{DY}}_{m,n,s,l}(t)$ are the path loss and path length of the l-th dynamic subpath as a function of time t, respectively. The precondition of Eq. (5) is that the numbers of static subpaths and dynamic subpaths do not change in a short period. When a short time interval τ elapses, Eq. (5) can be expressed by,

$$\begin{aligned} h_{m,n,s}(t+\tau) &= h^{\mathrm{ST}}_{m,n,s} + \sum_{l=1}^{L^{\mathrm{DY}}_{m,n}} \beta_{m,n,s,l}(t+\tau) e^{-j2\pi f_s \frac{d^{\mathrm{DY}}_{m,n,s,l}(t+\tau)}{c}} \\ &= h^{\mathrm{ST}}_{m,n,s} + \sum_{l=1}^{L^{\mathrm{DY}}_{m,n}} \beta_{m,n,s,l}(t+\tau) e^{-j2\pi f_s \frac{d^{\mathrm{DY}}_{m,n,s,l}(t)+v_{m,n,s,l}(t)\tau}{c}} \\ &= h^{\mathrm{ST}}_{m,n,s} + \sum_{l=1}^{L^{\mathrm{DY}}_{m,n}} \beta_{m,n,s,l}(t+\tau) e^{-j(2\pi f_s \frac{d^{\mathrm{DY}}_{m,n,s,l}(t)}{c} + \Delta\varphi_{m,n,s,l}(t))}, \end{aligned} \tag{6}$$

where $v_{m,n,s,l}(t)$ is the rate of change of the length of the l-th subpath at time t, and $\Delta\varphi_{m,n,s,l}(t)$ has the following form,

$$\Delta\varphi_{m,n,s,l}(t) = 2\pi f_s \frac{v_{m,n,s,l}(t)\tau}{c}, \tag{7}$$

The frequency change value can be obtained from Eq. (7), and the Doppler frequency shift at time t is as follows,

$$f_{m,n,s,l}(t) = \frac{\Delta\varphi_{m,n,s,l}(t)}{2\pi\tau} = f_s \frac{v_{m,n,s,l}(t)}{c}. \tag{8}$$

In an indoor wireless channel, a single moving object may affect many subpaths. Ignoring the NLOS scenario, only one of these subpaths is a primary reflection path, and the energy of the remaining multiple reflection paths is relatively low and can be ignored. Therefore, the rate of change of the path length inversely deduced by the Doppler frequency shift can reflect the size of the projection of the moving speed in a certain direction [21], and this direction is the normal direction of the ellipsoid where the transceiver end is the focus, and the current position is located.

3.1.2. Fresnel Zone Model

Fresnel zone model often appears together with Doppler spread model. Fresnel zone refers to an ellipsoid with the positions of the transmitter and receiver as the focus, as shown in Fig. 2. Point B_n on these ellipsoids satisfies the following distance constraints,

$$|d_{\mathrm{TBn}}| + |d_{\mathrm{BnR}}| - |d_{\mathrm{TR}}| = \frac{n\lambda}{2} \quad \text{for } n = 1, 2, \ldots, \tag{9}$$

where λ is the wavelength. The space between the $(n-1)$-th ellipsoid and the n-th ellipsoid is called the n-th Fresnel zone. The first Fresnel zone (FFZ) is the area enclosed by the first ellipsoid, which is the area that contributes the most to the received signal. The influence of the object in the FFZ on the wireless channel is mainly the diffraction phenomenon, and the reflection phenomenon is the main phenomenon in other Fresnel zones.

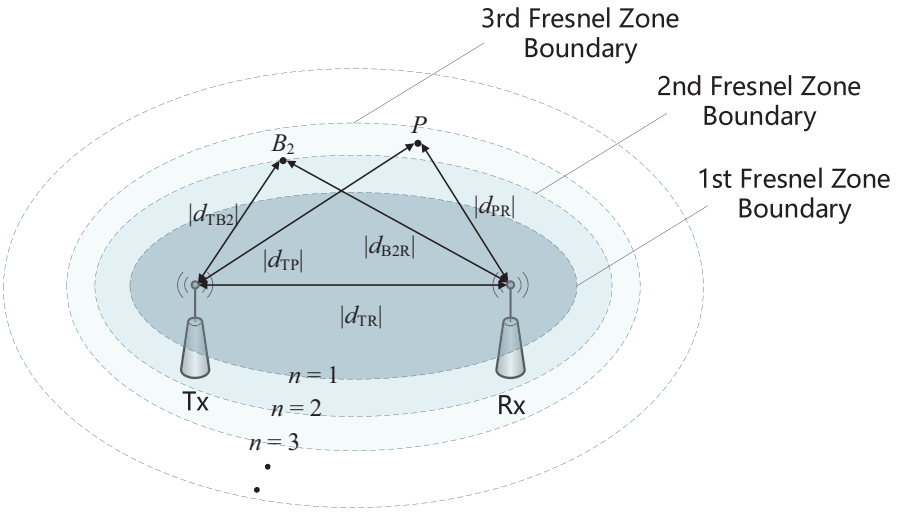

Figure 2. Fresnel zone is the area surrounded by ellipsoids with the transceiver end as the focus.

In the Fresnel zone except the FFZ, reflection phenomena dominate. Compared with LoS path, the path length of the reflection path is longer. Therefore, there is a certain phase difference between the reflection path and LoS path at the receiving end. Considering any reflection point P in space, without blocking the LoS path, the signal at the receiving end is the superposition of the LoS path and reflection path. The phase difference between the reflection path and LoS path can be expressed by the path length difference as follows,

$$\Delta\varphi = 2\pi\frac{\Delta d}{\lambda} = 2\pi\frac{|d_{\mathrm{TP}}| + |d_{\mathrm{PR}}| - |d_{\mathrm{TR}}|}{\lambda}, \tag{10}$$

When the reflection point is located at the boundary of the Fresnel zone, the phase of the reflection path lags behind the LoS path by $n\pi$. When n is an odd number, the two signals cancel in the opposite directions, and when n is an even number, the two signals are enhanced in the same direction. When the reflection point is located in the Fresnel zone. The closer the reflection point is to the boundary of the odd Fresnel zone, the closer the reflection signal and LoS path signal are to the opposite direction. On the contrary, the closer the two signals are to the same direction. And the amplitude of the superimposed signal is between the amplitude ranges on the boundary of two adjacent Fresnel zones at this point. Therefore, when a moving object crosses the boundaries of odd and even Fresnel zones, the received signal will show fluctuations in which troughs and peaks appear in sequence.

Since objects in the FFZ are closer to the receiver, the diffraction phenomenon of the objects dominates. In particular, when the object blocks the LoS path, the original LoS path will become the transmission path, and the diffraction phenomenon will be more obvious in the total received signal. The premise of the Fresnel integral used to analyze the diffraction phenomenon is that the incident

electromagnetic wave is a uniform plane wave, which is difficult to satisfy in the space-constrained indoor situation. But it still has a certain reference value. The work of Zhang et al. verifies that when an object vertically passes through the LoS path in a straight line in the FFZ, the normalized amplitude of the signal received by the receiving antenna exhibits a symmetrical "w" shape [22, 23].

The Fresnel zone model maps the movement of objects in the Fresnel zone to signal fluctuations and is often used for positioning and tracking [24, 21], as well as some fine-grained reciprocating motion detection near the transceivers [25, 23]. Combined with Doppler spread model mentioned earlier, more accurate velocity estimation can be achieved.

3.1.3. AoA and ToF

Among other model-based methods, the angle of arrival (AoA) and time of flight (ToF) models are also commonly used for localization and tracking [26, 17, 19]. AoA and ToF represent the angle and distance from the signal source to the receiver, respectively. The position of the signal source is determined by using the AoA or ToF of multiple receivers or using AoA and ToF at the same time. Space-Alternating Generalized Expectation-Maximization (SAGE) and Multiple Signal Classification (MUSIC) are two algorithms commonly used to estimate AoA and ToF. The SAGE algorithm is a deterministic parameter estimation algorithm. The delay and angle parameters are estimated by alternate iterative search methods. The goal of the iterative search is to make the likelihood function reach the maximum value, and each group of parameters converges to a stable value finally. Compared with the SAGE algorithm, the MUSIC algorithm is a classic high-resolution estimation algorithm. It is based on the eigen-decomposition class of the covariance matrix and constructs a subspace for parameter estimation. As shown in Fig. 3, the direction angle and time delay of the incoming wave can be obtained by searching for spectral peaks.

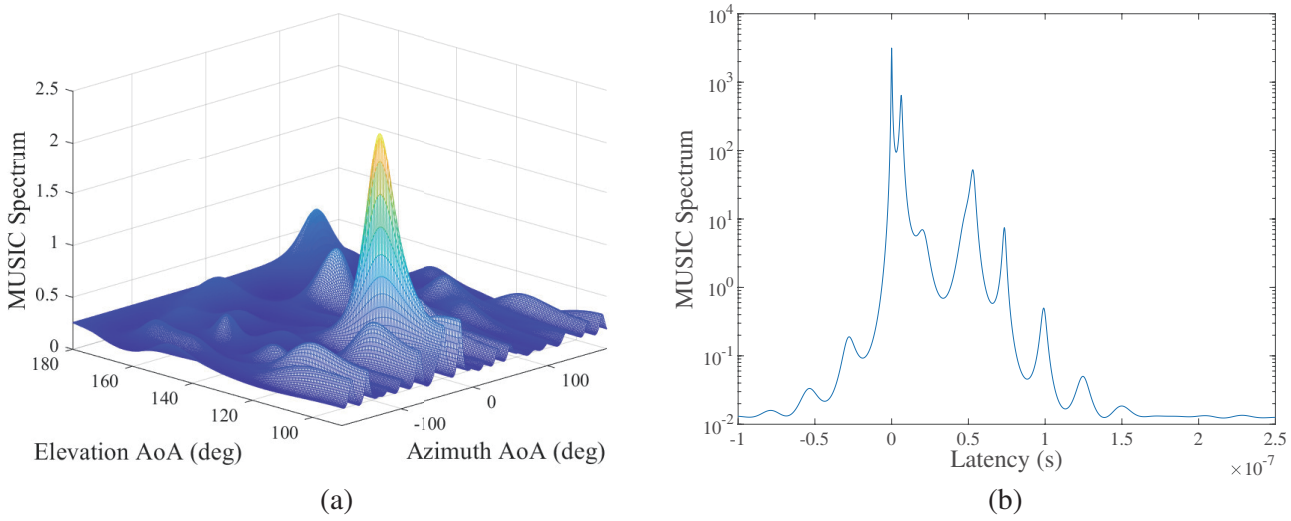

Figure 3. The MUSIC spectrum of AoA and ToF. (a) AoA estimation in the spatial domain. (b) ToF estimation in the frequency domain.

3.1.4. Others

ToF can also be roughly estimated by calculating the power delay profile (PDP) of the received signal. The PDP represents the strength of the received signal at different delays. The PDP can be obtained by performing an inverse fast Fourier transform (IFFT) on the signal in the subcarrier domain. The power spectral density (PSD) is a statistical model. According to the Wiener-khintchine theorem, the PDP is the Fourier transform of the signal autocorrelation function. It defines how the signal power is distributed with frequency. Autocorrelation and cross-correlation are also commonly used statistical models. Autocorrelation is the cross-correlation of a signal at different times, whereas cross-correlation

is a metric to calculate the similarity between two signals. Cross-correlation of different sets of measured CSI to determine similarity is often used in identification applications, while autocorrelation of CSI to observe its temporal similarity is often used in detection applications.

In addition, the short-time Fourier transform (STFT) is a transform commonly used for time-frequency domain analysis of received data. STFT can determine the frequency domain characteristics of time-varying signals in different time windows. Different from the traditional Fourier transform, STFT can reflect the time domain and frequency domain characteristics of the signal at the same time. Its essence is to window the signal and perform Fourier transform on the windowed signal, respectively. The window function is fixed, and the resolution is determined accordingly. Through the moving window function, the spectrum of the signal is taken segment by segment. The STFT $y(n, \omega)$ of a discrete signal $x(n)$ can be expressed as,

$$y(n, \omega) = \sum_{m} x(m) w(n - m) e^{-j\omega m}, \tag{11}$$

where $w(n)$ is a window function with a fixed window size.

3.2. Data-Based Methods

The achievements of AI in computer vision are obvious to all. Would it be helpful for sensing to replace image data in computer vision with CSI data in ICAS? The answer is positive. A large number of different types of measured data in sensing applications provide a natural data foundation for machine learning. Some classification problems can be solved by using machine learning to extract subtle differences between different datasets and learn commonalities within the same dataset. This provides a viable avenue for many sensing scenarios that cannot be accurately modeled.

The existing ICAS works involve numerous learning algorithms. Pattern recognition algorithms are commonly used in speech recognition, and optical character recognition (OCR) fields still attract much attention in ICAS. This type of algorithm achieves the purpose of recognition by finding special patterns in various types of data. The input of the pattern recognition algorithm is the preprocessed data features, so the extraction of the features directly determines the upper limit of the accuracy of the algorithm. Unlike pattern recognition, another direct method is to input the data set intactly into the deep learning algorithm and let the machine automatically learn the hidden features of various categories of data sets. Deep learning eliminates the manual feature extraction process and instead uses feedforward networks to learn autonomously from the data.

Deep learning is generally considered a type of pattern recognition algorithm, and the practice of inputting compressed data into deep neural networks for learning [27] also exists. But as far as the current CSI-based ICAS systems are concerned, the machine learning algorithms they use tend to fall into two categories: artificially extracting features and using non-deep neural network classifiers, inputting pre-compressed data or raw data into deep neural networks for learning. Therefore, the above-mentioned divisions of pattern recognition and deep learning are still representative and even more relevant.

3.2.1. Pattern Recognition

The support vector machine (SVM) [28] is a generalized linear classifier for binary classification of data. It maps the labeled feature vectors in the training set directly to points in space and then finds a hyperplane in this space that best separates these differently labeled points. The feature vectors of the data to be classified are mapped to both sides of the hyperplane through the same mapping to determine the corresponding category. Suppose that the training set is $\mathbf{D}$, $\mathbf{y}$, where $\mathbf{D} = \{\mathbf{D}_1, \mathbf{D}_2, \ldots, \mathbf{D}_N\}$, $\mathbf{D}_i = [d_1, d_2, \ldots, d_M]$ is the feature vector of the i-th data; $\mathbf{y} = \{y_1, y_2, \ldots, y_N\}$ is the learning target; and $y_i \in [-1, 1]$ is a binary variable, indicating that the i-th data corresponds to a negative class or a positive class. Map $\mathbf{D}$ directly into space $\mathbf{X} = \{(x_1, \ldots, x_M) | x_1, \ldots, x_M \in \mathbb{R}\}$. The form of the hyperplane that can separate datasets is as follows,

$$\mathbf{w}^T \mathbf{X} + b = 0 \quad \text{s.t. } y_i \left(\mathbf{w}^T \mathbf{D}_i + b\right) \geqslant 1, \tag{12}$$

where $\mathbf{w}$ and b are the normal vector and intercept of the hyperplane, respectively. For the linearly separable case, there are infinitely many such hyperplanes. The goal of the SVM is to find a hyperplane

with the largest shortest distance from the dataset to this hyperplane. For some linearly inseparable cases, the original data can be mapped to a higher-dimensional space by performing a nonlinear mapping, and different types of point sets in this high-dimensional space are linearly separable. However, this method increases the computational complexity exponentially. Kernel function is a method provided by the SVM that can maintain low complexity and solve linear inseparable problems.

The K-nearest neighbor (KNN) [29] is a supervised learning algorithm commonly used in classification and regression. The idea is that if most of the K most similar samples in the feature space of a sample belong to a certain category, then the sample also belongs to this category. The metric for judging this similarity is the distance between samples in the sample space. The smaller the distance is, the more similar the two samples are. For two samples $\mathbf{x} = (x_1, x_2, \ldots, x_N)$, $\mathbf{y} = (y_1, y_2, \ldots, y_N)$ containing N dimension features, the Minkowski distance is defined as follows,

$$D(\mathbf{x}, \mathbf{y}) = \left(\sum_{n=1}^{N} |x_n - y_n|^p \right)^{\frac{1}{p}}, \tag{13}$$

where p can be any value. When $p = 1$, it is called Manhattan distance. When $p = 2$, it is called Euclidean distance. When $p = \infty$, it is called the Chebyshev distance.

The K value, as the only hyperparameter in the KNN, directly affects the final effect. If the value of K is too small, the approximation error of learning will decrease, but the estimation error will increase, which is prone to overfitting. When the value of K is too large, the learning error will decrease, but the approximation error will increase, which is prone to underfitting. Generally, in applications, a smaller K value is taken first, and then the cross-validation method is used to finally determine the K value with the highest accuracy.

Another supervised learning algorithm commonly used in classification and regression in machine learning is the decision tree (DT) algorithm, which uses information entropy as a metric to construct a tree with the fastest drop in entropy. The metrics of information entropy include the information gain used in ID3 [30], the information gain rate used in C4.5 [31], and the Gini coefficient used in classification and regression tree (CART) [32]. In ICAS, branch nodes are constructed with data features and thresholds, and the categories of each group of data are used as leaf nodes. The best decision tree model is obtained through training. Bagging strategy works by sampling with replacement in the sample set and constructing a decision tree on all features for these sampled samples. After repeating multiple times to obtain multiple trees, the classification is achieved by voting. The random forest [33] is improved based on bagging strategy. After sampling with replacement, the features are also randomly sampled. The random forest can avoid overfitting more effectively while improving generalization ability.

The hidden Markov model (HMM) [34] is a statistical model that uses the time-dependent features of data to build state transition models. It has been widely used in various recognition applications and plays a significant role in speech recognition. In the ICAS classification problem, the observation sequence for each activity can be treated as a Markov model. The visible state of the HMM is the processed feature; the hidden state of the HMM is the decomposition of the activity in different periods; and the observation state of the HMM is the category after classifying the features. The change of the feature at adjacent moments corresponds to a state transition. Hidden states are derived from the observed features, after which the HMM identifies and classifies them by studying changes in different activity time domains.

3.2.2. Deep Learning

Deep learning is a machine learning method based on neural network algorithms. In 2006, Hinton et al. published an article in Science [35], proposing a solution to the gradient vanishing problem in deep neural network (DNN), which caused the upsurge of deep learning once again. Nowadays, there are many variants of deep learning algorithms on the original DNN algorithm, such as deep residual network (ResNet) [36], convolutional neural network (CNN) [37], recurrent neural network (RNN) [38], and long short-term memory (LSTM) [39].

The DNN contains an input layer, multiple hidden layers, and an output layer. The number of neurons in the input and output layers is equal to the length of the input sample data vector and the desired output vector, respectively. Too few neurons in the hidden layer will lead to underfitting, and

too many will lead to overfitting. Therefore, the optimal number of neurons in the hidden layer needs to be obtained through continuous verification. The layers are fully connected and forward propagated through linear combinations and activation functions. The output of the neuron in l-th layer can be expressed as follows,

$$\mathbf{a}^l = \sigma\left(\mathbf{z}^l\right) = \sigma\left(\mathbf{W}^l\mathbf{a}^{l-1} + \mathbf{b}^l\right), \tag{14}$$

where $\mathbf{z}^l$ is the linear combination of the outputs of the neurons in the $(l-1)$-th layer, which can be regarded as the input of the neurons in the l-th layer. $\mathbf{W}^l$ and $\mathbf{b}^l$ are the weight coefficient matrix and paranoid matrix of the $(l-1)$-th layer to the l-th layer, respectively, and $\sigma(\cdot)$ is the activation function. There are various activation functions. Hidden layers usually use the Tanh function and the ReLU function. The output layer selects different activation functions according to different application scenarios. For regression problems, the identity function can be used to output the data that is input to the output layer as it is. For binary classification problems, the sigmoid function can be used. For multivariate classification problems, the softmax function can be used.

The training process of DNN is to find the best $\mathbf{W}$ and $\mathbf{b}$ matrices so that the output of the training sample set and the output of DNN maintain a high consistency. The input vector of the sample set is input into the initial DNN for training through backpropagation, and $\mathbf{W}$ and $\mathbf{b}$ are optimized using the loss function as the metric. Commonly used loss functions are mean square error and cross-entropy error. The optimization method is gradient descent, including batch gradient descent, mini-batch gradient descent, stochastic gradient descent (SGD), and stochastic gradient descent with momentum (SGDM).

However, it has been proved that the fully connected DNN requires more computing power with the increase of the number of layers, and it is easy to fall into the local optimum due to the phenomenon of overfitting. This greatly limits the performance of neural networks. In the work of ICAS, CNN is often used instead of DNN for classification and recognition [16, 27, 40–46]. CNN considers the spatial distribution of samples and introduces the concept of the convolution kernel. CNN replaces hidden layers in DNN with convolutional layers and pooling layers. In the convolutional layer, the relationship between the output of the convolutional layer and the previous layer is determined by the convolution kernel, stride, and activation function. The pooling layer samples the output of the convolutional layer to reduce the dimension and compress the features. Pooling operations include max pooling, average pooling, and L2 pooling. Finally, the result is output through a fully connected layer. CNN can be seen as a special kind of DNN. It greatly reduces the network parameters, thus effectively avoiding the network falling into local optimum.

DNN and CNN assume that all inputs are independent, while many of the samples in the recognition of combined actions and gestures in ICAS research are temporally correlated. At this point, RNN with memory can be used to deal with such problems. RNN consists of an input layer, hidden layers, and an output layer. The hidden layer in RNN depends not only on the current input but also on the value of the previous hidden layer, which can be expressed as follows,

$$\mathbf{a}_t = \sigma(\mathbf{U}\mathbf{x}_t + \mathbf{W}\mathbf{a}_{t-1} + \mathbf{b}), \tag{15}$$

where $\mathbf{a}_t$ and $\mathbf{a}_{t-1}$ are the output of the hidden layer at the current moment and the previous moment, respectively; $\mathbf{x}_t$ is the input of the current hidden layer; $\mathbf{U}$ and $\mathbf{W}$ are the weight coefficient matrixes of the input at the current moment and the output of the hidden layer at the previous moment, respectively; $\mathbf{b}$ is the paranoid matrix; and $\sigma()$ is the activation function. Besides, $\mathbf{U}$, $\mathbf{W}$, $\mathbf{b}$ are the same at any time.

The RNN processes data sequentially in such a way that the current output is influenced by previous input data. The RNN can be regarded as a neural network that propagates on time, and its depth is the length of time. However, this simple RNN structure is prone to gradient vanishing and gradient explosion in the temporal dimension. The LSTM is by far the most used solution. LSTM selects important information to retain and forgets unimportant information through the gate structure. As an improved RNN, as shown in Fig. 4, the LSTM can perform better in longer sequences.

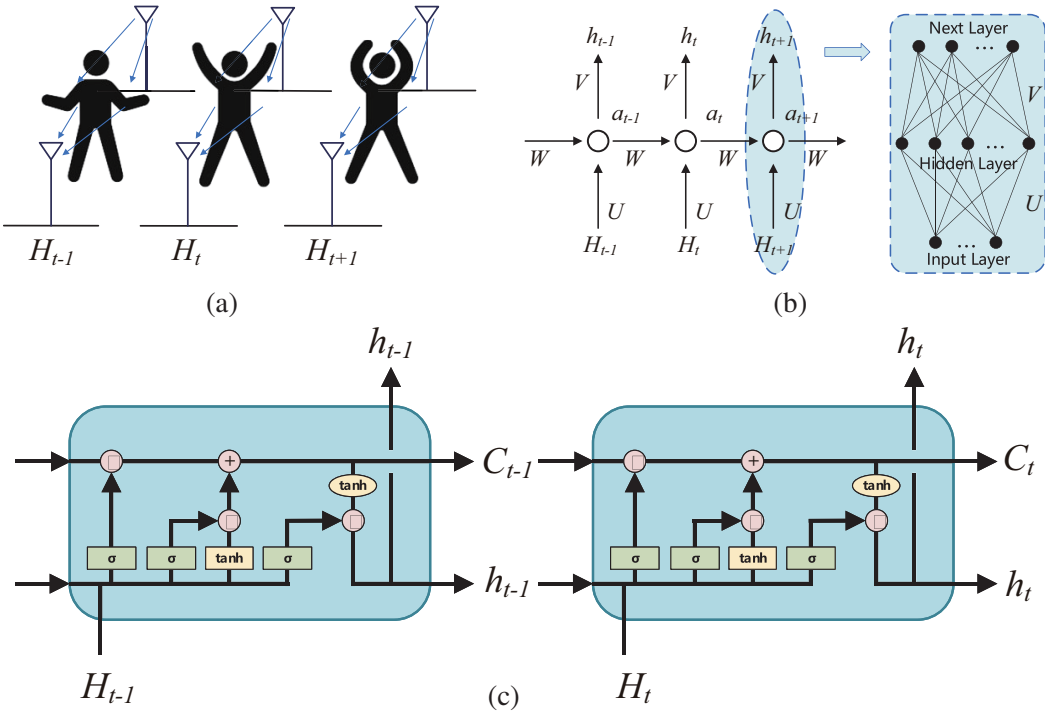

Figure 4. The temporal features of time-related motion can be learned by inputting the collected data into the RNN or LSTM. (a) Sequence CSI from time-dependent combined motion. (b) Input the collected CSI sequences to the input layer of the RNN in sequence. (c) Input the collected CSI sequences to the input layer of the LSTM in sequence.

3.3. Model-Data Hybrid-Driven Methods

Model-data hybrid-driven methods effectively combine models from model-based methods with learning algorithms from data-based methods. This method first builds a physical model of the problem to be solved to obtain rough features of the problem, then machine learning algorithms are used to refine them. Comparing the advantages and inevitable shortcomings of pure model-based and data-based methods shows that these two methods are complementary in many aspects. Model-data hybrid-driven methods combine their advantages and break down their respective limitations.

The advantages and disadvantages of these three methods and future development trends are summarized as follows:

Model-based methods start from the characteristics of electromagnetic waves and build special physical models for different scenarios and applications. The original information is transformed into a form that can reflect environmental information through signal processing. Compared with the other two methods, model-based methods do not require a large amount of data to be collected for training, and the algorithm is simple enough to achieve real-time sensing. Model-based methods are widely applicable to detection and estimation applications. However, each model in the model-based approach is designed for only one application, so it is difficult to extend. And there are a lot of problems that cannot be modeled in actual scenes, such as complex action recognition. For most recognition applications, model-based methods cannot accurately distinguish between different classes through modeling. The accuracy of model-based methods is limited by accurate measurement and environmental dynamic information extraction, so more effective noise filtering, multipath elimination, and dynamic information extraction algorithms will become the focus of research.

On the other hand, data-based methods start from the measurement data and adopt different learning methods according to the requirement to learn the features hidden in the data, to achieve the purpose of prediction, identification, or classification. The learned data can be CSI raw data or feature vectors extracted from CSI. Compared with the other two methods, data-based methods do not

depend on the spatial context and do not require extensive signal processing. Data-based methods are more suitable for dealing with some problems in recognition and detection scenarios. But it also has many flaws. High recognition accuracy often requires a lot of training costs. Then, various activities at various locations in the space are reflected differently on the receiving end, and it is impossible for data-based methods to take them all into account. Many data-based methods only identify and classify special activities in a specific location, and basically cannot be further extended to the entire space or to achieve simultaneous detection of various fine-grained information. The accuracy of data-based methods greatly depends on the prediction accuracy of machine learning algorithms. In addition, most motions in human sensing are time-related. In the future, machine learning algorithms that consider time characteristics will become the focus of research.

As mentioned earlier, model-based and data-based methods are complementary in many ways. Model-data hybrid-driven methods learn the results obtained based on the model, which can achieve more accurate estimation and detection results, and even realize the sensing function that could not be achieved before. Model-data hybrid-driven methods preprocess the input of the algorithm in the data-based methods. Compared with the previously unprocessed data, the processed data can already reflect some spatial characteristics. Therefore, the training cost is greatly reduced, which provides the conditions for giving play to the full potential of machine learning. This hybrid method can often achieve higher accuracy than using models or learning algorithms alone, and also provides more ideas and solutions for ICAS work. Model-data hybrid-driven methods not only require the selection of the proper model and learning algorithm but also need to consider how to effectively combine the data processed by the model with the data-based algorithm, which is also the main research direction of model data hybrid-driven methods in the future.

4. APPLICATIONS OF CSI-BASED ICAS

Different application scenarios in CSI-based sensing are summarized in this section and divided into three categories: detection, estimation, and recognition. According to the above-mentioned three methods, scenarios, and specific applications in each scenario, the researched articles are summarized in Table 1.

In addition, this section focuses on representative applications in each scenario, and the devices, protocols, models, signal processing methods, machine learning algorithms, and performance used in these works are summarized according to Fig. 5.

4.1. Detection

In detection scenarios, the CSI-based sensing system needs to detect specific actions or objects in the environment. This application scenario requires that the system is only sensitive to a certain change in the environment and does not need to react to other things happening in the environment.

4.1.1. Human Detection

The simplest and most common detection application is human detection, which judges whether there is a person in the environment. It is suitable for home scenarios and security fields. It is also the basis for all human sensing. Some human detection works detect people in motion and judge whether there is a person by comparing the dynamic and static features of the environment [47, 48]. R-TTWD in [49] can realize dynamic human detection through the walls by calculating the first-order difference mean of different subcarriers as features. This work of detecting people by distinguishing between static and dynamic environments is essentially Motion Detection. WiDetect in [50] implements a calibration-free motion detection system by exploiting the difference in the autocorrelation properties of CSI in static and dynamic environments.

However, it is not accurate enough to judge the existence of people only by detecting the dynamic changes of the environment, and people are not always in motion in the environment. Omni-PHD in [51] detects the human through the fingerprint-based method and can judge the azimuth angle of the human relative to the AP. In addition, DeMan in [52] detects stationary people by using a bandpass filter and sine fitting to capture the environmental changes caused by people's breathing while achieving dynamic human detection.

Table 1. Summary of existing WLAN ICAS.

Type	Scenario	Application	Reference
Model-Based	Detection	Human Detection	[52, 50]
		Full Detection	[17, 53]
		Smoke Detection	[54]
		LoS Detection	[55, 56]
	Estimation	Walking Direction Estimation	[57]
		Human Counting	[58–60]
		Speed Estimation	[53]
		Indoor Positioning and Tracking	[26, 17, 19–21, 24, 61]
		Imaging	[15, 62, 63]
		Respiration Monitoring	[25, 64, 65, 52]
	Recognition	Gesture Recognition	[66, 17, 67, 19]
		Passing Direction Identification	[60, 68]
Data-Based	Detection	Human Detection	[47–49, 51]
		Full Detection	[69]
	Estimation	Human Counting	[70–72]
		Indoor Positioning	[48, 73, 74, 16]
	Recognition	Keystroke Recognition	[75, 76]
		Human Identification	[77, 78, 42]
		Gesture Recognition	[44, 79, 80, 81]
		Human Activity Recognition	[82–84, 73, 41, 42, 27] [43, 16, 85–90]
Model-Data Hybrid-Driven	Detection	Full Detection	[91, 92]
	Estimation	Human Counting	[93–95, 40, 46, 18]
		Speed Estimation	[94, 91]
		Respiration Monitoring	[93, 96]
	Recognition	Human Identification	[97, 98]
		Gesture Recognition	[99–101]
		Speech Recognition	[102]
		Human Activity Recognition	[95, 45, 103]
		Passing Direction Identification	[94, 40]

Remarks: Human detection research work is summarized in Table 2. The detection system targeting the dynamic human body is essentially a dynamic detection system, whose key point is how to extract the different characteristics in the CSI of the static and dynamic environments. Once features with high discrimination are found, simple binary classifiers or threshold-based detection methods can be used for human detection. However, the impact of the static human body on the environment is weak, and capturing such weak changes requires the establishment of high-precision physical models or the direct use of fingerprint-based methods.

4.1.2. Fall Detection

Fall detection is designed for special family scenarios. For the elderly who are physically fragile, falls are more harmful to the body than adults and children. Real-time monitoring of fall accidents in a non-contact way through Wi-Fi devices at home, and responding in the first time, can gain more rescue

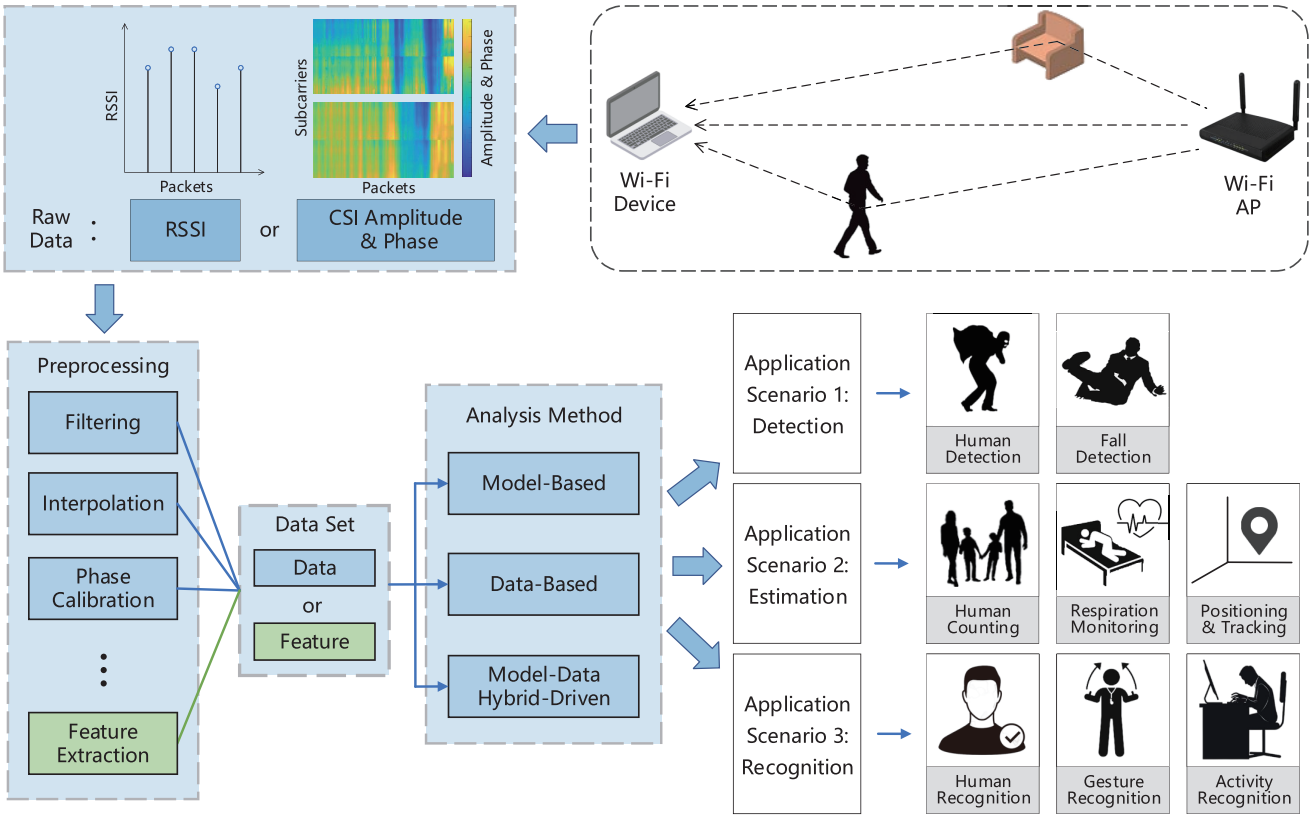

Figure 5. WLAN-Based ICAS technologies and their applications.

time for the elderly at home alone.

WiFall in [69] is a data-based fall detection system. It detects abnormal activities by calculating local outlier factor, then extracts features to distinguish falling activities from walking, sitting, and standing activities using SVM or Random Forest. RT-Fall in [51] adopts the same strategy as WiFall, first detecting abnormal activities and then extracting features using SVM to identify fall activities. DeFall in [91] and WiSpeed in [53] detect fall activity by estimating the velocity and acceleration of motion in the environment over a period of time. The difference is that DeFall uses the SLN-DTW algorithm, and WiSpeed adopts a model-based method, which estimates speed by calculating autocorrelation and then detects fall activity.

Remarks: Fall detection research work is summarized in Table 3. The fall detection system is often required to determine the time-interval of the fall movement to extract its features. Next, to distinguish fall movement from other fall-like motions, such as walking and sitting, the traditional approach is to extract features and classify them. Most of the classifiers used are binary classifiers or threshold-based discriminators. In addition, some velocity estimation work can achieve fall detection by predicting the velocity and acceleration of the movement.

4.1.3. Other Detection Applications

Other detection applications are summarized in Table 4. Smokey in [54], a smoking activity monitoring system in public places, can monitor smoking activities in the environment. Smokey proposes a motion acquisition method based on foreground detection and filters counterfeit foregrounds using temporal and frequency correlation. Finally, the unique breathing pattern of smoking is used to determine whether each compound action is a smoking action.

The dependence of indoor positioning and other sensing applications on LoS path makes LoS/NLoS detection a fundamental primitive. Besides, LoS/NLoS detection is beneficial to communication

Table 2. Summary of detection scenarios: Human detection.

Reference	Other Function	Base Signal	Stationary or Dynamic Human	Signal Processing	Method	Performance
DeMan [52]	/	Amplitude & Phase	Both	Hampel Filter BPF, LMS	Correlation Sinusoidal Model Nelder-Mead	Accuracy: 94.82%/93.33%, dynamic/static
WiDetect [50]	/	-	Dynamic	/	Autocorrelation Thresholding	Accuracy: $\geqslant$ 83%
PADS [47]	/	Amplitude & Phase	Dynamic	Phase Calibration Hampel Filter	Thresholding SVM	Accuracy: 97%
Pilot [48]	Positioning	Amplitude	Dynamic	/	Maximum a Priori Probability Fingerprint-Based Methods	Accuracy: 90%, Positioning Accuracy: 90%
R-TTWD [49]	/	Amplitude	Dynamic	Hampel Filter Interpolation Wavelet Filter Normalization PCA	SVM	True Positive/ True Negative: 99%
Omni-PHD [51]	Azimuth Distinction	Amplitude & Phase	Both	/	Fingerprint-Based Methods EMD	Accuracy: $\geqslant$ 91%, Azimuth Distinction: $\geqslant$ 75%

Notice: All the studies in this table are implemented on IEEE 802.11n protocol by using the Intel 5300 NIC.

link setting adjustment for high throughput and reliable communication. LiFi in [56] exploits the characteristic that the received signal envelope exhibits Rician distribution under LOS propagation and Rayleigh distribution under NLOS conditions. LiFi extracts the envelope distribution features Rician-K factor and skewness of the received signal. Then set thresholds to detect LoS and NLoS paths. PhaseU in [55] uses spatial diversity and frequency diversity to distinguish LoS and NLoS by thresholding based on LiFi.

4.1.4. Summary

In detection applications, both model-based and data-based methods are popular. Commonly used models include correlation characteristics and time-frequency domain analysis [17, 50, 53–56, 91, 92]. Most of the learning algorithms are light-weight classifiers such as SVM and random forest [47, 49, 69, 92] or use EMD and DTW [51, 91] for binary classification.

4.2. Estimation

In estimation scenarios, the system needs to make judgments against uncertainties in the environment. Unlike the system in the detection scene that only needs to make binary judgments, the possible results of the problem in the estimation scene are mostly continuous. For example, speed estimation requires the system to give an exact number as the output result [53, 91, 94].

Table 3. Summary of detection scenarios: Fall detection.

Reference	Other Function	Base Signal	Transmission Rate	Signal Processing	Method	Performance
WiSpeed [53]	Speed Estimation	-	150 0Hz	Median Filter l_1 Trend Filter	Autocorrelation Peak Detection	Accuracy: 95% Speed Mean Error: 4.85%
WiFall [69]	/	Amplitude	100 Hz	WMAF, CSI Aggregation Local Outlier Factor, SVD	SVM or Random Forest	Accuracy: 90%/94% False Alarm Rate: 15%/13% (SVM/RF)
DeFall [91]	/	Amplitude	1500 Hz	/	Autocorrelation SLN-DTW DBA, WGSS Thresholding	Accuracy: 96% False Alarm Rate: 1.47%
RT-Fall [92]	/	Amplitude & Phase	100 Hz	BPF, Phase Calibration Interpolation	Phase Difference STFT, SVM Thresholding	True Positive Rate: 91%, True Negative Rate: 92%

Notice: All the studies in this table are implemented on IEEE 802.11n protocol by using the Intel 5300 NIC.

Table 4. Summary of detection scenarios: Other detection applications.

Reference	Application	Platform	Signal Processing	Method	Performance
Smokey [54]	Smoking Detection	Intel 5300 802.11n	Interpolation	Foreground Detection Temporal/ Frequency Correlation Peak Detection	True Positive Rate: 97.6% False Positive Rate: 0.8%
PhaseU [55]	LOS/NLoS Identification	Intel 5300 802.11n	Hampel Filter Phase Calibration Antenna Selection	Phase Difference Thresholding	Detection Rate: 94%/80%, (static/mobile)
LiFi [56]	LOS/NLoS Identification	Intel 5300 802.11n	Phase Calibration Normalization IFFT	Rician & Rayleigh Distribution Thresholding	Detection Rate: 90.42%, False Alarm Rate: 9.34%

4.2.1. Human Counting

Human counting requires the system to judge the number of people in the current environment. FCC in [58] proposes a metric, the percentage of nonzero elements (PEM), in the dilated CSI Matrix. The grey Verhulst model is used to fit the monotonic relationship between PEM and the number of people

to estimate. [94] adopts a nonlinear fitting method to model the monotonic relationship between CSI features and the number of people in the environment. In addition, it constructs the crowd density contour curve according to the CSI amplitude distribution on the subcarriers under different crowd densities. Then the crowd density is determined by calculating the EMD distance. Meanwhile, this system uses the relative phase and total harmonic distortion algorithm to estimate the moving speed of the person. WiCount in [71] directly inputs the preprocessed CSI amplitude and phase into the BP neural network for learning. The system predicts the number of people by fitting the correlation between the number of people and CSI by determining the parameters of the neural network.

In addition to using fitting to achieve human counting, it can also be considered a classification problem. [70, 18] both use feature selection and linear discriminant analysis (LDA) to estimate the number of people. [72] preprocesses the CSI amplitude and extracts its features, then inputs it into a multi-class SVM model using error-correcting output codes (ECOC) for classification.

People flow counting is a special case of human counting, which requires the system to estimate the number of people passing through an area. Wi-Count in [60] takes the CSI phase as the research goal and uses the FastICA algorithm to separate the influence of a single person to achieve the purpose of counting. Door-Monitor in [40] changes the research target to the CSI phase difference of the two antennas based on Wi-Count and replaces the FastICA algorithm with STFT analysis and CNN. WiFlowCount in [46] takes the CSI conjugate product of two antennas as the research object. People flow counting is realized through four stages: flow detection, subflow detection, subflow size estimation, and continuous flow counting. The detection of the first two stages is achieved by threshold detection. The last stage inputs the STFT spectrogram of the substream into CNN for learning to estimate the number of people.

Remarks: Human counting research work is summarized in Table 5. The influence of the number of people on the channel is reflected in multiple reflection paths and fading. There are two types of traditional research methods. The first method is to find the monotonic relationship between the number of people in the environment and some characteristics of CSI. Then build a special model to fit this monotonic relationship. Another approach is to treat population estimation as a classification problem, where the class is the number of people. Moreover, because people flow counting studies the number of people at a particular location, the transmitting and receiving devices tend to be short distances. At this point, more granular analysis methods are needed to extract the different effects caused by different numbers of people.

4.2.2. Respiration Monitoring

Respiration monitoring requires the system to capture the subtle changes in the environment caused by human respiration and to estimate the respiration rate and cycle, which is of great significance in the medical field. The Wi-Sleep system in [25, 96] places the transceiver devices on the two ends of the lying down human chest to capture human breathing. It performs a series of preprocessing on the collected raw CSI data and then selects the subcarrier CFR sequence with the strongest periodicity to track respiration. At the same time, the DBSCAN clustering method is used to detect whether the sleeping posture has changed, and three sets of transceiver devices in different positions are used to judge the sleeping posture. PhaseBeat in [64] uses the MUSIC algorithm to estimate the respiratory rate and the FFT-based method to estimate the heart rate. TR-BREATH in [93] estimates the breathing rate by calculating the time-reversal resonating strength and analyzing it using the MUSIC algorithm. TR-BREATH uses clustering by affinity propagation, likelihood assignment, and cluster merging to estimate the breathing rate of multiple people simultaneously, and cluster-based estimation is used to estimate the number of people. Since the effect of respiration on the signal is weak, respiration detection is extremely dependent on the static environment.

Remarks: Respiratory monitoring research work is summarized in Table 6. Due to the strong periodicity of human respiration, the influence of respiration on the signal is concentrated in the frequency domain. For single-person monitoring scenarios, the respiration rate can be estimated using the respiration-induced sine-like periodic pattern changes in CSI amplitude or phase. For multi-person monitoring scenarios, it is necessary to use the differences in respiration rates of different people to achieve monitoring. In practical environments, the effect of respiration is very weak and difficult

Table 5. Summary of estimation scenarios: Human counting.

Reference	Application	Platform	Base Signal	Signal Processing	Method	Performance
FCC [58]	HC	Intel 5300 802.11n	Amplitude	/	Rician Fading Grey Verhulst Model, PEM	Errors $\leqslant$ 2: 98%/70% (indoor/outdoor)
[70]	HC	Intel 5300 802.11n	Amplitude	Normalization	Davies-Bouldin index, LDA	Errors $\leqslant$ 2: 81%
[18]	HC	Ettus USRP N210 802.11b	Amplitude & Phase	DFT	PPMCC, LDA	Accuracy: CSI 72%/69% RSSI 70%/57% (Office/Meeting)
[94]	HC, Density Estimation Speed & Direction Derivation	Intel 5300 802.11n	Amplitude & Phase	LPF Subcarrier Selection	Non-Linear Regression, EMD Phase Difference THD, FFT	Accuracy: $\geqslant$ 90%
WiCount [71]	HC	Intel 5300 802.11n	Amplitude & Phase	LPF, Phase Calibration WMAF	BPNN	Accuracy: 82.3%
[72]	HC	Nexmon CSI Extractor 802.11n	Amplitude	Hampel Filter LPF	ECOC-SVM	Accuracy: 92%/97% (Elevator/Office)
Wi-Count [60]	PFC, PDD	Intel 5300 802.11n	Phase	K-means Savitzky-Golay Filter	FastICA Peak Detection	PDD: 95% PFC: 92%
Door-Monitor [40]	PFC, PDD	Intel 5300 802.11n	Phase Difference	K-means Median filter Savitzky-Golay Filter	STFT, CNN	PDD: 95.2% PFC: 94.5%
WiFlow-Count [46]	PFC	Intel 5300 802.11n	CSI Conjugate Product	BPF, PCA Binarization Gauss Filter	Thresholding STFT, CNN	Accuracy: 96.7%/94.3% (People Flow $\leqslant$ 4/6)

Notice: HC represents Human Counting. PFC represents People Flow Counting. PDD represents Passing Direction Detection.

to detect, and most work is extremely dependent on static environments. Monitoring respiration in complex dynamic environments remains a difficult problem.

4.2.3. Indoor Positioning and Tracking

Indoor positioning and tracking is also a specific application of estimation scenarios, which requires the system to judge and continuously track the position of the person. Widar in [24] is a model-based indoor positioning and tracking system. The Doppler model is utilized to locate and track by extracting the path length change rate. Widar 2.0 in [61] reduces the two links in Widar to one and adds analysis of ToF and AoA. Both Wits in [26] and IndoTrack in [21] are two-link indoor positioning and tracking systems that track by estimating Doppler shifts. Regarding the initial positioning, Wits uses the ToF of the two links, and IndoTrack uses the AoA to determine. The former parameter estimation algorithm

Table 6. Summary of estimation scenarios: Respiration monitoring.

Reference	Other Function	Platform	Transmission Rate	Signal Processing	Method	Performance
Wi-Sleep [25, 96]	Posture Change Detection	Intel 5300 802.11n	20 Hz	Hampel Filter Interpolation Wavelet Filter Subcarrier Selection	SVD, STFT Nelder-Mead Method Kolmogorov-Smirnov test, DBSCAN	Tracking Respiration: FPR & FNR $\leqslant$ 27% Posture Change Detection: $\geqslant$ 80% Apnea: $\geqslant$ 85%
Phase-Beat [64]	/	Intel 5300 802.11n	400 Hz	Hampel Filter Subcarrier Selection DWT, FFT	Phase Difference Peak Detection root-MUSIC	Maximum Estimation Error: 0.85/10 bpm (Breathing/Heart)
TR-BREATH [93]	Human Counting	Wi-Fi NIC 802.11n	10 Hz	CSI Calibration	TRRS, root-MUSIC Time-Reversal K-means, SVM Likelihood Assignment	Breathing Rate LoS: 99.56%/98.65% NLoS: 99.37%/97.3% (Single/Multi-Person)

is maximum likelihood while the latter is MUSIC algorithm. DFLAR in [73] converts CSI of multiple channels into radio images and extracts color and texture features from radio images. The optimized deep image features are then extracted from the original image features using a sparse auto-encoder (SAE) network. Finally, the deep features are input into the softmax regression algorithm to estimate the position of the person, and at the same time, it can recognize the activity of the person.

Remarks: Indoor positioning and tracking research work is summarized in Table 7. The implementation schemes of CSI-based indoor positioning are mainly divided into two categories: triangulation and fingerprint-based methods. Triangulation mainly involves the estimation of parameters such as AoA and ToF and needs to use channel parameter estimation algorithms such as MUSIC, SAGE, and maximum likelihood. The estimation accuracy of AoA and ToF is limited by the number of antennas and bandwidth, respectively. Better super-resolution parameter estimation algorithms have become a mainstream research direction. Fingerprint-Based Methods achieve indoor positioning by building a fingerprint library and matching it with the collected data. This method usually achieves higher accuracy but is inflexible and extremely sensitive to environmental changes. In addition to real-time positioning, Doppler and Fresnel zone models can be used to predict motion trajectories to assist in tracking.

4.2.4. Summary

Model-based methods are widely used in estimation applications. For example, Doppler and Fresnel zone models [57, 24, 61, 21, 25], as well as MUSIC algorithm [21, 15, 64, 93], are commonly used in respiration monitoring, direction estimation, and indoor positioning. ToF is also commonly used for indoor positioning [26, 17, 19, 61]. Time-frequency domain analysis of data is widely used in various applications [57, 17–19, 24, 64, 65, 40, 46, 94, 96]. Data-based methods can solve classification problems for estimation applications, or directly transform estimation problems into classification problems. Both pattern recognition algorithms [18, 93, 96, 70, 72, 74] and deep learning algorithms [40, 46, 73, 16, 71] have applications.

4.3. Recognition

In recognition scenarios, the target to be recognized must be clearly defined first. Then the system needs to identify the target existing in the current environment based on the collected signals. There are many applications in the recognition scene. Only three main applications are selected here for detailed introduction.

Table 7. Summary of estimation scenarios: Indoor positioning and tracking.

Reference	Other Function	Platform	Transceiver Device Configuration	Signal Processing	Method	Performance
Widar [24]	Speed Estimation	Intel 5300 802.11n	1Tx-2Rx(3) (1.6, 1.6) m 2000 Hz	Subcarrier Selection BPF, PCA STFT	Doppler Spread PLCR Extraction Cross-Correlation Thresholding Least Fitting Error	Median Error: 0.38 m Median Relative Velocity Error: 13%
Widar2.0 [61]	/	Intel 5300 802.11n	1Tx-1Rx(3) -, 1000 Hz	GPM, Conjugate Multiplication Kalman Filter	Doppler Spread SAGE, AoA, ToF	Average Error: 0.75 m
Wits [26]	/	-	1Tx-2Rx(3) (2.8, 3.2) m /(4, 4) m, -	Phase Calibration Static Elimination	Doppler Spread Maximum Likelihood, ToF	Mean Error: 0.41 m
IndoTrack [21]	/	Intel 5300 802.11n	1Tx-2Rx(3) (6, 6) m 200 Hz	Conjugate Multiplication Static Elimination	Doppler Spread Fresnel Zone MUSIC, AoA	Median Error: 0.35 m
DFLAR [73]	Activity Recognition	Intel 5300 802.11n	1Tx-1Rx(3) 13 m, 500 Hz	Median Filter GLCM, Phase Calibration Gabor Filter	Softmax Regression Algorithm, Sparse Auto-Encoder Network	Accuracy: $\geqslant$ 90% (5 Activities at 12 Locations)

***Notice*:** Transceiver device configuration includes the number of transceiver devices, the number of antennas in the device, the distance between the transceiver devices, and the transmission rate.

4.3.1. Human Recognition

One of the applications that we are exposed to the most daily is human recognition. Face recognition, fingerprint recognition, and even iris recognition, genetic recognition are its implementation methods. CSI-based human recognition is achieved in a contactless manner by studying the influence of people on Wi-Fi signals in the environment.

WiFi-ID in [77] selects the time and frequency domain features from the CSI of the walking state to construct fingerprints by the ReliefF feature selection algorithm. Then the Sparse Approximation based Classification face recognition method proposed in [104] is used for recognition. FreeSense in [78] performs discrete wavelet transform compression on CSI sequences, and a DTW-based KNN classifier is used to identify gait. The learning algorithms used by WiWho in [97] and Wi-FiU in [98] are DT and SVM, respectively. EfficientFi in [42] proposes the first Wi-Fi sensing framework to support IoT cloud. Robust CSI features are extracted using CNN in Wi-Fi and compressed into a quantized vector using the CSI codebook for transmission. In the cloud server, the vector is restored and reconstructed as the original CSI data, which is input into a recognition classifier composed of fully connected layers. EfficientFi reduces communication overhead while achieving high-precision recognition.

Remarks: Human recognition research work is summarized in Table 8. CSI-based human recognition studies the different influences of different people on CSI. Among them, identifying people by studying the walking characteristics of different people is the most common. This research method is also called gait recognition. However, both gait recognition and general human recognition need to filter out the identity-independent quantities in CSI to obtain pure identity features. Therefore, as

Table 8. Summary of recognition scenarios: Human recognition.

Reference	Other Function	Transceiver Device Configuration	Signal Processing	Method	Performance
Wi-Fi-ID [77]	/	1Tx(3)-1Rx(3) 2 m, 800 Hz	Silence Removal CWT, ReliefF	Sparse Approximation based Classification	Accuracy: 93% ∼ 77% (From 2 ∼ 6 People)
FreeSense [78]	/	1Tx(2)-1Rx(3) 2.5 m, 1000 Hz	LPF, DWT, PCA	DTW, KNN	Accuracy: 94.5% ∼ 88.9% (From 2 ∼ 6 People)
EfficientFi [42]	Activity Recognition	1Tx(3)-1Rx(3) 3 m, 500 Hz	Feature Compression and Reconstruction	CNN	Activity Recognition Accuracy: 98% Human-ID: 89.5%
WiWho [97]	/	1Tx(3)-1Rx(3) 7.2/8.2/10 m, 100 Hz	/	IFFT, DTW Peak Detection Decision Tree	Accuracy: 92% ∼ 80% (From 2 ∼ 6 People)
Wi-FiU [98]	/	1Tx(2)-1Rx(3) 1.6 m, 2500 Hz	PCA, Frequency Domain Denoising	STFT, SVM	Accuracy: 79.28% From 7 People

Notice: All the studies in this table are implemented on IEEE 802.11n protocol by using the Intel 5300 NIC.

shown in Table 8, most human recognition work has higher requirements on the distance between the transceiver and the human recognition area. This greatly limits the application prospects of CSI-based human recognition systems. Next, reducing the sampling rate of the recognition system should also be considered in subsequent human recognition research work.

4.3.2. Gesture Recognition

CSI-based gesture recognition provides a very convenient interaction method for human-computer interaction. Hand-level gesture recognition focuses on the movement of the entire hand and tracking the trajectory of the hand. Therefore, hand-level gesture recognition can also be referred to as hand tracking. WiG in [81] uses Birge-Massart filter to filter and local outlier factor (LOF) based anomaly detection algorithm to detect the anomaly segment on the collected CSI. Finally, the extracted features are input to the SVM for classification. WiTrace in [66] uses the Doppler spread model to implement 1D and 2D hand tracking. WiTrace designs a heuristic algorithm to achieve the purpose of static vector elimination and uses linear regression to remove the sampling frequency offset (SFO) and packet detection delay (PDD) offset. Afterward, the phase difference between the sub-carriers is used to estimate the path length of the dynamic path, and the initial position estimation is performed in two steps. Finally, Kalman filtering based on a 2D continous Wiener process acceleration (CWPA) model is used to correct the error caused by the continuous noise of the system.

The target of finger-level gesture recognition is finger movements. This application recognizes more complex gestures and even sign language, so finger-level gesture recognition can be referred to as sign language recognition. SignFi in [44] is a data-based sign language recognition system. This system uses multiple linear regression on the CSI phase to remove the frequency offset and then inputs the

magnitude and phase together into a 9-layer CNN network for learning and classification. SGDM is used in SignFi to update the weights and biases. The WiFinger in [100] uses pattern matching to recognize gestures based on CSI, which uses Multidimensional Dynamic Time Warping (MD-DTW) for similarity calculation after CSI preprocessing. WiFinger in [101] is an American sign language (ASL) recognition system that uses a DTW-based KNN classifier to recognize ASL.

Table 9. Summary of recognition scenarios: Gesture recognition.

Reference	Specific Application	Platform	Transceiver Device Configuration	Signal Processing	Method	Performance
WiG [81]	Hand Tracking	Intel 5300 802.11n	1Tx(3)-1Rx(2) -, 100 Hz	Birge-Massart Filter, Lof Anomaly Detection	SVM	4 Gesture Accuracy: 92%/88% (LoS/NLoS)
WiTrace [66]	Hand Tracking	Ettus USRP N210 802.11g	1Tx(1)-2Rx(1) (0.5, 0.5) m 100Hz	Hampel Filter Normalization LPF, Static Elimination Linear Regression Kalman Filter	Thresholding STFT, Doppler Spread	Initial Position Estimation: 6.23 cm, Average Tracking Errors: 1.46/2.09 cm (1D/2D Tracking)
SignFi [44]	Sign Language Recognition	Intel 5300 802.11n	1Tx(1)-1Rx(3) 1.3/2.3 m, 200 Hz	Multiple Linear Regression	CNN, SGDM	276 Gestures Accuracy: 98% (Lab/Home) 94% (Lab+Home)
WiFinger [100]	Sign Language Recognition	Intel 5300 802.11n	1Tx(-)-1Rx(-) -, 20 Hz	BPF, DWT	Thresholding MD-DTW IFFT, Pattern Matching	8 Gestures Accuracy: 93%
WiFinger [101]	Sign Language Recognition	Intel 5300 802.11n	1Tx(1)-1Rx(3) 0.5 m, 2000 Hz	Hampel Filter LPF, WMA DWT	Autocorrelation Thresholding KNN, DTW	Finger-Grained Accuracy: 90.4%/82.67% (9/90 Digits)

Remarks: Gesture recognition research work is summarized in Table 9. The influence of hand movements in wireless channels is very weak, and there are certain similarities between different gestures. Therefore, it is difficult to obtain discriminative features using model-based methods. Research using purely machine learning algorithms usually has a precondition that the gesture occurs at a fixed location close to the AP. Although this approach is beneficial to achieve high-precision recognition, it sacrifices the robustness and flexibility of the system. Considering that vision-based or wearable device-based solutions can realize gesture recognition more easily and accurately, future research work on CSI-based gesture recognition should give full play to its own advantages and avoid disadvantages.

4.3.3. Activity Recognition

Human activity recognition aims to identify daily activities. The human activities that most of the works focus on are simple basic activities such as walking, running, sitting, and falling. We refer to the system that recognizes this type of activity as atomic activity recognition system. In contrast to atomic

activity, a system that recognizes some combined actions, such as exercise, laundry, watching TV, and other daily activities, can be called the combined activity recognition system.

CARM in [82, 90] proposes a CSI-speed model to relate the frequency power variation of CSI to human movement speed. The CSI-activity model correlates the movement speed of different parts of the human body with specific human activities. Finally, the HMM is used to identify eight kinds of human daily activities. ReWiS in [27] uses SVD to compress the CSI sequence and inputs it into the few-shot learning of the ProtroNet for learning. WiSDAR in [45] and [43] use deep learning architectures combining CNNs with LSTMs. [43] proposes to adopt a matching network with enhanced channel state information (MatNet-eCSI) to facilitate one-shot learning human activity recognition. [85] proposes attention based bi-directional long short-term memory (ABLSTM) based on LSTM. WiAct in [86] designs a novel adaptive activity cutting algorithm (AACA) based on the difference in signal variance between the action and non-action parts. After that, the magnitude of the conjugate multiplication of the two-link CFRs is used as the feature for activity recognition, and a three-layer ELM is used for learning.

E-eyes in [84] use EMD techniques to measure the distribution of measured CSI and profiles to identify in situ activity. Moving activity is identified using MD-DTW to align CSI measurements with known activities in the profile. E-eyes can identify 13 daily activities such as eating, cooking, sleeping, bathing, and washing according to different detection areas. WiFit in [103] and SEARE in [89] are two exercise monitoring systems that identify an exercise action by its specific impact on CSI and perform continuous counting and monitoring.

Remarks: Activity recognition research work is summarized in Table 10. Due to the complexity of human daily activities and the location of occurrence is not fixed, it is difficult to establish an accurate model to analyze the movements in daily activities. In this case, the data-based approach for machines to learn the characteristics of different activities and the differences between them has been proven to be a feasible approach. Besides, compared to pattern recognition algorithms, deep learning can extract features from data more efficiently. In the future, with the development of deep learning, a large number of novel deep neural networks will be introduced for the problem of activity recognition that cannot be modeled.

4.3.4. Other Recognition Applications

In recognition scenarios, in addition to the above three main applications, there is some special but very meaningful work. These other recognition applications are summarized in Table 11. WiKey in [75, 76] recognizes keystrokes by placing the keyboard 30 cm from the receiver. WiKey achieves an average keystroke recognition accuracy of 96.4% for 37 input keys using a DTW-based KNN classifier. WiHear in [102] utilizes MIMO beamforming technology and directional antennas to locate and focus the mouth to achieve speech recognition. The system uses Multi-Cluster/Class Feature Selection (MCFS) to extract features. Finally, generalized least squares and DTW are used to obtain the similarity of the current signal to the profile for classification.

4.3.5. Summary

Since recognition involves the comparison and learning of different categories of datasets, data-based methods are widely used in recognition scenarios. For example, common algorithms in pattern recognition algorithms are SVM [83, 99, 98, 103], KNN [83, 99], HMM [82, 99, 90], DT [97, 99], and DTW-based KNN or SVM [78, 75, 88, 76, 101]. There are also studies using DTW-based pattern matching [84, 102, 100, 79, 89, 80], and ELM classifiers [86]. In deep learning, commonly used algorithms are CNN [40–42, 27, 16, 44], LSTM [85], and the learning architecture combining CNN and LSTM [43, 45]. Model-based methods in recognition applications are often used to assist in extracting features for machine learning, so time-frequency domain analysis is often used [97, 102, 94, 40, 98, 100]. Fresnel zone and Doppler models [99, 45, 103] are often used to guide AP deployment or assist in feature extraction.

Table 10. Summary of recognition scenarios: Activity recognition.

Reference	Type of Activity	Standard	Transmission Rate	Signal Processing	Method	Performance
CARM [82, 90]	Atomic	Intel 5300 802.11n	2500 Hz	PCA, EMA DWT	Thresholding HMM	8 Activities Accuracy: 96%
ReWiS [27]	Atomic	Nexmon Tool 802.11ac	100 Hz	Normalization SVD	ProtoNet, FSL CNN	4 Activities Accuracy: 98.85%
WiSDAR [45]	Atomic	Intel 5300 802.11n	500 Hz	LPF, PCA	Fresnel Zone STFT, LSTM Thresholding, CNN	8 Activities Accuracy: 96% FPR $\leqslant$ 10%
[43]	Atomic	Intel 5300 802.11n	1000 Hz	Correlation	EWMA-Recursion MatNet, CNN BLSTM	7 Activities Accuracy: 86.8%/93.4% (One/Five-Shot)
[85]	Atomic	Intel 5300 802.11n	500 Hz	-	ABLSTM, ADAM	8 Activities Accuracy: 96.7%
WiAct [86]	Atomic	Intel 5300 802.11n	1000 Hz	LPF Median Filter Sliding Variance	Thresholding ELM	10 Activities Accuracy: 94.2%
E-eyes [84]	Combined	Intel 5300 802.11n	20 Hz	LPF, MCS Index Filter	EMD, MD-DTW Pattern Matching Non-Profiling Clustering	13 Activities Accuracy: 90%/94% (Single/Multiple Wi-Fi Device)
WiFit [103]	Exercise	Intel 5300 802.11n	200 Hz	-	Doppler Spread MUSIC, DTW SVM	3 Activities Accuracy: 95.8%, Repetition Counting: 99%
SEARE [89]	Exercise	Intel 5300 802.11n	50 Hz	LPF, PCA Median Filter	FoD-FFT, DTW Thresholding	4 Activities Accuracy: 97.8%/91.2% (LoS/NLoS)

5. OPEN PROBLEMS AND FUTURE DIRECTIONS

With the evolution of communication technology, in the future, massive Machine-Type Communication (mMTC) scenarios will open up a convenient path for the intelligent IoT in the field of communication, and the construction of the indoor intelligent environment will also progress with the continuous development of WLAN. We have introduced the research progress of CSI-based ICAS in various indoor scenarios, then the notable challenges and possible future development directions are present here.

Table 11. Summary of recognition scenarios: Other recognition applications.

Reference	Application	Platform	Transmission Rate	Signal Processing	Method	Performance
WiKey [75, 76]	Keystrokes Recognition	Intel 5300 802.11n	2500 Hz	LPF, PCA, DWT	DTW, KNN	Detection Rate: 97.5% Recognition: 96.4%
WiHear [102]	Speech Recognition	Intel 5300/ Ettus USRP N210 802.11n	100 Hz	BPF, DWT	FFT, MCFS DTW, SRDA	Accuracy: 91%/74% (1/3 Person)

5.1. Open Problems

5.1.1. Robustness and Universality

The robustness of an algorithm refers to its anti-noise ability. In the ICAS system, this robustness can be understood as the sensitivity of the WLAN system to the system and environment configuration or some special parameters in the algorithm. Most of the work introduced before places the transmitting and receiving devices in the middle of an clear environment, and some special sensing systems even require fixed transceiver positions. However, the location of Wi-Fi in home scenarios is uncertain, and in office scenarios, APs are mostly located on the ceiling. Responsive, dynamic changes in the environment relative to the location of the transmitter and receiver are also a thorny problem. The influence of the relative positions of the transmitter, the receiver, and the person is the first problem that the ICAS system needs to overcome. Secondly, the adaptability of the system in the new unknown environment and the accurate judgment of the new unknown data category are also crucial. These challenges determine whether the system can move from the lab to the masses. Finally, the system needs to be robust to more specific issues such as network configuration, movement speed, and number of multipaths. The fundamental way to overcome this challenge and guarantee the generalization ability of the system is to find suitable models, extract the most effective features and use machine learning algorithms rationally.

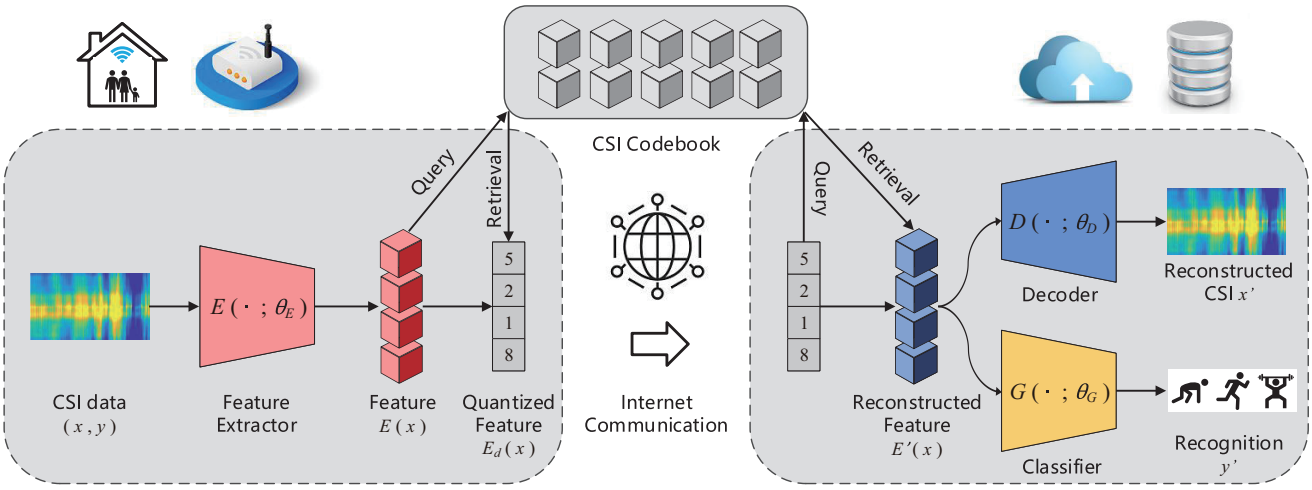

Figure 6. The design of the proposed EfficientFi framework.

5.1.2. Reasonable Coexistence of Communication and Sensing

The CSI-based ICAS system requires Wi-Fi devices to continuously transmit and collect measurement signals at high frequencies. The overhead incurred can only be borne by the communication performance. At the same time, these measurement signals will have an effect on other radio devices in the environment. In order to reduce the overhead caused by the high sampling frequencies of the sensing function, the DeFall system in [91] adds a pre-judgment module. The module monitors the environment at a lower sampling frequency (30 Hz). When the module detects dynamic changes in the environment, the system will perform fall detection at a high sampling frequency. Under the premise of maintaining a high sampling rate (500 Hz), the EfficientFi system in [42] performs feature extraction and codebook compression on the CSI on the Wi-Fi side and then transmits it to the cloud server, as shown in Fig. 6. The identification takes place in the cloud server, which also reduces communication overhead.

5.2. Future Directions

5.2.1. More thorough and Comprehensive Sensing

At present, the ICAS research is still in the exploratory stage. Although it is regarded as an achievable vision by 6G and next-generation Wi-Fi, its key technologies, indicators and specific implementation forms are still incomplete. Therefore, a more thorough and comprehensive study on sensing technology is definitely the focus of future research work.

First of all, most of the current ICAS technology research is based on single-person scenarios. However, in complex indoor multi-person scenarios, the system is required to sense multiple objects simultaneously. Moreover, the mutual interference between different objects is an inevitable problem in multi-person scenarios. Therefore, the ICAS technologies for single-person scenarios are usually difficult to be applied to multi-person scenarios directly. Using special signal processing algorithms to separate multi-person information into single-person information is the method adopted by most of the ICAS works that consider multi-person scenarios. Such as, the root-MUSIC [105] is used by [64, 93] to estimate the respiratory rate of multiple people. The influence of different people on the signal is modeled as a linear superposition of different components in [60] when detecting the simultaneous entry and exit of multiple people, and ICA [106] is used to separate each component. Through this method, the system for single-person scenarios can be generalized to multi-person scenarios. However, the accuracy of the system is also affected by the accuracy of single-person information extraction. In [95], the correlation matrix of CSI amplitude and phase is used as an auxiliary feature for discrimination to reduce the activity detection error caused by the multi-person dynamic environment.

Secondly, among the articles surveyed above, there are only a few works [93, 102, 54, 100, 67, 19, 91, 53, 89] that investigate the system performance in NLoS scenarios. But it is very necessary for the actual environment. At the same time, these works only compare the sensing performance in NLoS scenarios and do not propose substantive solutions. Based on 802.11ac, an outdoor human detection is investigated in [107], and then a multi-station method is used to extend to NLoS scenarios [108], with an accuracy rate of 99.58%. It is well known that the channel bias introduced by NLOS has serious effects on ToA-based positioning systems. To solve this difficulty, a Bayesian approach is developed in [109], where the joint distribution of all channel bias values is modeled with a multivariate GMM to account for the channel correlations. In addition, diffraction-based and reflection-based millimeter-wave radar sensing systems in NLoS scenarios are respectively studied in [110, 111]. In the future, the ICAS system can also be developed toward channel model analysis and human sensing solutions in NLoS scenarios.

Last but not least, the human sensing achieved by current ICAS research work is only the first step in constructing an intelligent environment. To realize the true interconnection of all things in indoor scenarios, the sensing of all things is an indispensable part. There should be corresponding sensing solutions for people, objects, or pets. Next, the research on multi-device interconnection and cross-layer collaboration is also crucial. The research in this paper is mainly aimed at the CSI-based human sensing technology, and only involves the information of the Wi-Fi physical layer. However, the IoE requires the use of various available devices and technologies to collect information on any object that needs to be monitored, connected, and interacted. In the future, the sensing device of ICAS can be extended to multiple Wi-Fi devices or introduce a variety of different sensors to realize the ICAS. It is also possible

to improve the entire ICAS system architecture by combining information from the upper layers of the physical layer [112–114].

5.2.2. Waveform Design for Both Communication and Sensing

The OFDM technology widely used in WLAN and cellular systems are susceptible to severe Doppler spread. Waveforms are susceptible to inter-carrier interference. Furthermore, OFDM is not as important in sensing as it is in communication due to its ambiguous autocorrelation and high peak-to-average power ratio (PAPR) [7]. The ICAS waveform design aims to find a dual-function waveform that takes both communication and sensing performance into consideration.

In the radar communication system, to solve the hardware pressure caused by the high peak-average ratio of the traditional OFDM waveform, the CE-OFDM based on OFDM in [115] is a new waveform with high speed, anti-fading, and constant envelope for the RadCom system. One power minimization-based robust radar waveform design criteria are proposed in [116]. The radar transmitted power can be efficiently reduced by exploiting the communication signals scattered off the target at the radar receiver. In addition, to achieve a favorable performance trade-off between radar and communications, [117] derives a novel communications metric as a function of both subcarrier powers and forwarded control information.

Modifying Frequency Modulated Continuous Wave (FMCW) technology, which is commonly used in high-precision radar ranging, to enable communication has also become a research trend in ICAS waveform design. FMCW has the characteristics of simple waveform and low PAPR, but cannot modulate and transmit data. The phased coded FMCW proposed in [118] combines the communication capabilities of PMCW and the sensing capabilities of FMCW. And PC-FMCW proposed by [119] combines phase encoding with linear frequency modulation (LFM) continuous wave. And effectively increase the interference suppression capability of the system. The PARC framework is extended to FMCW operation in [120] which maximizes both data throughput and energy on target. In addition to phase encoding, [121, 122] encodes information onto FMCW radar signals by amplitude modulation to realize the transmission of communication data.

However, since the above designs are implemented by adding sensing or communication capabilities to OFDM or FMCW, they cannot get rid of the constraints of the original framework. As a result, these works cannot be well adapted to indoor scenarios with multiple access and high-speed data transmission requirements, which are the main usage scenarios of ICAS in the future. In recent years, a novel two-dimensional modulation scheme, the Orthogonal Time-Frequency Space (OTFS) modulation technique [123], has been regarded as a promising candidate for high-mobility communications. OTFS modulates information in the delay-Doppler (DD) domain rather than in the time-frequency (TF) domain of classical OFDM modulation [124]. It applies a 2D inverse symplectic finite Fourier transform (ISFFT) to the baseband data to transform it into the time-frequency domain and obtains the transmitted signal through the Heisenberg transform. OTFS modulation can convert a time-varying channel into a two-dimensional time-invariant channel in the DD domain, then exploit its properties. In scenarios of high Doppler spread, OTFS can achieve better performance than OFDM [125]. Therefore, OTFS is a feasible candidate in both sensing and future high-speed mobile communication scenarios.

5.2.3. Mutual Assistance of Communication and Sensing

To achieve a comprehensive integrated design of communication and sensing, communication and sensing need to be fully integrated in all aspects, such as spectrum resources, hardware equipment, waveform design, and networking collaboration. Communication and sensing functions are expected to complement each other for mutual benefit and become endogenous functions of the system.

With the continuous evolution of mobile communications, the deployment of small cells envisioned in the future Beyond 5G (B5G) and 6G systems offers unprecedented possibilities for ICAS [126]. In particular, 6G systems consider perception as a key endogenous function. In the future, B5G and 6G cellular systems will be able to provide ultra-high bandwidth and more abundant spectrum resources, which provide more possibilities for high-resolution sensing. Besides, with the increase in the density of cellular systems, future sensing will no longer be limited to indoor scenarios. And technologies such as beamforming and massive MIMO greatly improve not only the performance of the communication

system but also the sensing range and accuracy. Meanwhile, the combination of higher access numbers and transmission rates combined with the AI introduced in 6G makes real-time sensing easier. The introduction of new waveforms, such as the current 5GNR waveform [127], also provides convenience for implementing ICAS [128].

Conversely, sensing can also assist communication in many ways. In vehicle-to-infrastructure (V2I) communications, sensing-assisted beam prediction has gradually become a research direction [129–132]. Sensing can also improve user privacy security. [133] proposes to detect and track eavesdroppers and send jamming signals to the eavesdropper while the user is communicating to improve the secrecy performance of the communication. Furthermore, a sensing-aided inter-cell interference coordination (ICIC) technique is proposed in [134] to assist UAV resource block allocation. To avoid the strong interference problem caused by the UAV's LoS communication characteristics. At present, most of the research on sensing-assisted communication is in outdoor scenarios such as vehicle communication and UAV communication. The gains brought by next-generation WLAN and B5G for indoor sensing, such as indoor positioning and tracking and trajectory prediction, will provide reliable user information for communication systems.

6. CONCLUSIONS

A comprehensive survey on the CSI-based ICAS has been provided. First, the sensing techniques using the CSI of the IEEE 802.11 physical layer are described in detail. Next, the commercial devices and tools that can be used to collect CSI are presented, and their corresponding protocol standards are also given. Then, several commonly used physical models and machine learning algorithms are introduced, and their respective advantages and disadvantages in ICAS applications are analyzed. Moreover, the progress in various research aspects of CSI-based sensing combined with specific application examples is reviewed. Finally, the main challenges for future work in ICAS are outlined. We hope that this survey can help systematize existing ICAS works, inspire future directions, and stimulate wide applications.

ACKNOWLEDGMENT

This work was supported in part by the National Key R&D Program of China under Grant 2020YFB1804901, the National Natural Science Foundation of China under Grants 62001101 and 61960206006, the Science Foundation of Jiangsu Province of China under Grant BK20200349, the Southeast University-China Mobile Research Institute Joint Innovation Center, and the Keysight's University Research Collaborations program.

REFERENCES

1. You, X. H., C.-X. Wang, J. Huang, X. Q. Gao, Z. C. Zhang, M. Wang, Y. M. Huang, C. Zhang, Y. X. Jiang, J. H. Wang, et al., "Towards 6G wireless communication networks: Vision, enabling technologies, and new paradigm shifts," *Sci. China — Inf. Sci.*, Vol. 64, No. 1, 1–74, 2021.
2. HUAWEI, *Huawei Wi-Fi 6 (802.11ax) Technology White Paper*, 2019.
3. Viswanath, S. K., C. Yuen, W. Tushar, W.-T. Li, C.-K. Wen, K. Hu, C. Chen, and X. Liu, "System design of the internet of things for residential smart grid," *IEEE Wirel. Commun.*, Vol. 23, No. 5, 90–98, 2016.
4. Pokhrel, S. R., H. L. Vu, and A. L. Cricenti, "Adaptive admission control for IoT applications in home WiFi networks," *IEEE Trans. Mobile Comput.*, Vol. 19, No. 12, 2731–2742, 2019.
5. Chen, Q. H. and Y.-H. Zhu, "Scheduling channel access based on target wake time mechanism in 802.11ax WLANs," *IEEE Trans. Wireless Commun.*, Vol. 20, No. 3, 1529–1543, 2020.
6. He, Y., Y. Chen, Y. Hu, and B. Zeng, "WiFi vision: Sensing, recognition, and detection with commodity MIMO-OFDM WiFi," *IEEE Internet Things J.*, Vol. 7, No. 9, 8296–8317, 2020.
7. Paul, B., A. R. Chiriyath, and D. W. Bliss, "Survey of RF communications and sensing convergence research," *IEEE Access*, Vol. 5, 252–270, 2016.

8. Wang, Z. J., K. K. Jiang, Y. S. Hou, W. W. Dou, C. M. Zhang, Z. H. Huang, and Y. J. Guo, "A survey on human behavior recognition using channel state information," *IEEE Access*, Vol. 7, 155986–156024, 2019.
9. Ma, Y. S., G. Zhou, and S. Q. Wang, "WiFi sensing with channel state information: A survey," *ACM Comput. Surv.*, Vol. 52, No. 3, 1–36, 2019.
10. Rochim, A. F., B. Harijadi, Y. P. Purbanugraha, S. Fuad, and K. A. Nugroho, "Performance comparison of wireless protocol IEEE 802.11ax vs 802.11ac," *Proc. Int. Conf. Smart Technol. Appl. (ICoSTA)*, 1–5, 2020.
11. Nurchis, M. and B. Bellalta, "Target wake time: Scheduled access in IEEE 802.11ax WLANs," *IEEE Wirel. Commun.*, Vol. 26, No. 2, 142–150, 2019.
12. Deng, D.-J., Y.-P. Lin, X. Yang, J. Zhu, Y.-B. Li, J. Luo, and K.-C. Chen, "IEEE 802.11ax: Highly efficient WLANs for intelligent information infrastructure," *IEEE Commun. Mag.*, Vol. 55, No. 12, 52–59, 2017.
13. Halperin, D., W. J. Hu, A. Sheth, and D. Wetherall, "Tool release: Gathering 802.11 n traces with channel state information," *ACM SIGCOMM Comp. Commun. Rev.*, Vol. 41, No. 1, 53–53, 2011.
14. Xie, Y. X., Z. J. Li, and M. Li, "Precise power delay profiling with commodity Wi-Fi," *IEEE. Trans. Mob. Comput.*, Vol. 18, No. 6, 1342–1355, 2018.
15. Adib, F. and D. Katabi, "See through walls with Wi-Fi!," *Proc. ACM SIGCOMM Conf. SIGCOMM*, 75–86, 2013.
16. Wang, F., J. W. Feng, Y. L. Zhao, X. B. Zhang, S. Y. Zhang, and J. S. Han, "Joint activity recognition and indoor localization with Wi-Fi fingerprints," *IEEE Access*, Vol. 7, 80058–80068, 2019.
17. Adib, F., Z. Kabelac, D. Katabi, and R. C. Miller, "3D tracking via body radio reflections," *Proc. NSDI*, 317–329, 2014.
18. Cianca, E., D. S. Mauro, and D. D. Simone, "Radios as sensors," *IEEE Internet Things J.*, Vol. 4, No. 2, 363–373, 2016.
19. Adib, F., Z. Kabelac, and D. Katabi, "Multi-person localization via RF body reflections," *Proc. 12th USENIX Conf. Netw. Syst. Des. Implementation*, 279–292, 2015.
20. Xiong, J., K. Sundaresan, and K. Jamieson, "ToneTrack: Leveraging frequency-agile radios for time-based indoor wireless localization," *Proc. 21st Annu. Int. Conf. Mobile Comput. Netw.*, 537–548, 2015.
21. Li, X., D. Q. Zhang, Q. Lv, J. Xiong, S. J. Li, Y. Zhang, and H. Mei, "IndoTrack: Device-free indoor human tracking with commodity Wi-Fi," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, Vol. 1, No. 3, 1–22, 2017.
22. Zhang, F. S., D. Q. Zhang, J. Xiong, H. Wang, K. Niu, B. H. Jin, and Y. X. Wang, "From fresnel diffraction model to fine-grained human respiration sensing with commodity Wi-Fi devices," *Proc. ACM Interact. Mobile Wearable Ubiquitous Technol.*, Vol. 2, No. 1, 53, 2018.
23. Zhang, D. Q., F. S. Zhang, D. Wu, J. Xiong, and K. Niu, "Fresnel zone based theories for contactless sensing," *Contactless Hum. Activity Anal.*, 145–164, 2021.
24. Qian, K., C. S. Wu, Z. Yang, Y. H. Liu, and K. Jamieson, "Widar: Decimeter-level passive tracking via velocity monitoring with commodity Wi-Fi," *Proc. 18th ACM Int. Symp. Mobile Ad Hoc Netw. Comput.*, 1–10, 2017.
25. Liu, X. F., J. N. Cao, S. J. Tang, and J. Q. Wen, "Wi-Sleep: Contactless sleep monitoring via Wi-Fi signals," *Proc. IEEE Real-Time Syst. Symp.*, 346–355, 2014.
26. Tian, L. P., L. Q. Chen, Z. M. Xu, and Z. Chen, "Wits: An efficient Wi-Fi based indoor positioning and tracking system," *Remote Sens.*, Vol. 14, No. 1, 19, 2022.
27. Bahadori, N., J. Ashdown, and F. Restuccia, "ReWiS: Reliable Wi-Fi sensing through few-shot multi-antenna multi-receiver CSI learning," *arXiv preprint arXiv:2201.00869*, 2022.
28. Cortes, C. and V Vapnik, "Support-vector networks," *Mach. Learn.*, Vol. 20, No. 3, 273–297, 1995.

29. Cover, T. and P. Hart, "Nearest neighbor pattern classification," *IEEE Trans. Inf. Theory*, Vol. 13, No. 1, 21–27, 1967.
30. Quinlan, J. R., "Induction of decision trees," *Mach. Learn.*, Vol. 1, No. 1, 81–106, 1986.
31. Quinlan, J. R., *C4. 5: Programs for Machine Learning*, Elsevier, 2014.
32. Breiman, L., J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees*, Routledge, 2017.
33. Breiman, L., "Random forests," *Mach. Learn.*, Vol. 45, No. 1, 5–32, 2001.
34. Baum, L. E., T. Petrie, G. Soules, and N. Weiss, "A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains," *The Annals of Mathematical Statistics*, Vol. 41, No. 1, 164–171, 1970.
35. Hinton, G. E. and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, Vol. 313, No. 5786, 504–507, 2006.
36. He, K. M., X. Y. Zhang, S. Q. Ren, and J. Sun, "Deep residual learning for image recognition," *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 770–778, 2016.
37. LeCun, Y., B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Backpropagation applied to handwritten zip code recognition," *Neural Comput.*, Vol. 1, No. 4, 541–551, 1989.
38. Hopfield, J. J., "Neural networks and physical systems with emergent collective computational abilities," *Proc. Natl. Acad. Sci.*, Vol. 79, No. 8, 2554–2558, 1982.
39. Hochreiter, S. and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, Vol. 9, No. 8, 1735–1780, 1997.
40. Yang, Y. N., J. N. Cao, X. L. Liu, and X. F. Liu, "Door-monitor: Counting in-and-out visitors with COTS Wi-Fi devices," *IEEE Internet Things J.*, Vol. 7, No. 3, 1704–1717, 2019.
41. Yadav, S. K., S. Sai, A. Gundewar, H. Rathore, K. Tiwari, H. M. Pandey, and M. Mathur, "CSITime: Privacy-preserving human activity recognition using Wi-Fi channel state information," *Neural Netw.*, Vol. 146, 11–21, 2022.
42. Yang, J. F., X. Y. Chen, H. Zou, D. Z. Wang, Q. W. Xu, and L. H. Xie, "EfficientFi: Towards large-scale lightweight Wi-Fi sensing via CSI compression," *IEEE Internet Things J.*, 2022.
43. Shi, Z. G., J. Zhang, R. Y. D. Xu, and Q. Q. Cheng, "Environment-robust device-free human activity recognition with channel-state-information enhancement and one-shot learning," *IEEE. Trans. Mob. Comput.*, Vol. 21, No. 2, 540–554, 2020.
44. Ma, Y. S., G. Zhou, S. Q. Wang, H. Y. Zhao, and W. Jung, "SignFi: Sign language recognition using Wi-Fi," *Proc. ACM Interact. Mobile Wearable Ubiquitous Technol.*, Vol. 2, No. 1, 1–21, 2018.
45. Wang, F. X., W. Gong, and J. C. Liu, "On spatial diversity in Wi-Fi-based human activity recognition: A deep learning-based approach," *IEEE Internet Things J.*, Vol. 6, No. 2, 2035–2047, 2018.
46. Zhou, R., Z. Y. Gong, X. Lu, and Y. Fu, "WiFlowCount: Device-free people flow counting by exploiting Doppler effect in commodity Wi-Fi," *IEEE Syst. J.*, Vol. 14, No. 4, 4919–4930, 2020.
47. Qian, K., C. S. Wu, Z. Yang, Y. H. Liu, and Z. M. Zhou, "PADS: Passive detection of moving targets with dynamic speed using PHY layer information," *Proc. IEEE ICPADS*, 1–8, 2014.
48. Xiao, J., K. S. Wu, Y. W. Yi, L. Wang, and L. M. Ni, "Pilot: Passive device-free indoor localization using channel state information," *Proc. IEEE 33rd Int. Conf. Distrib. Comput. Syst. (ICDCS)*, 236–245, 2013.
49. Zhu, H., F. Xiao, L. J. Sun, R. C. Wang, and P. L. Yang, "R-TTWD: Robust device-free through-the-wall detection of moving human with Wi-Fi," *IEEE J. Sel. Areas Commun.*, Vol. 35, No. 5, 1090–1103, 2017.
50. Zhang, F., C. Chen, B. B. Wang, H. -Q. Lai, Y. Han, and K. J. R. Liu, "Widetect: A robust and low-complexity wireless motion detector," *IEEE ICASSP*, 6398–6402, 2018.
51. Zhou, Z. M., Z. Yang, C. S. Wu, L. F. Shangguan, and Y. H. Liu, "Towards omnidirectional passive human detection," *Proc. IEEE INFOCOM*, 3057–3065, 2013.

52. Wu, C. S., Z. Yang, Z. M. Zhou, X. F. Liu, Y. H. Liu, and J. N. Cao, "Non-invasive detection of moving and stationary human with Wi-Fi," *IEEE J. Sel. Areas Commun.*, Vol. 33, No. 11, 2329–2342, 2015.
53. Zhang, F., C. Chen, B. B. Wang, and K. J. R. Liu, "WiSpeed: A statistical electromagnetic approach for device-free indoor speed estimation," *IEEE Internet Things J.*, Vol. 5, No. 3, 2163–2177, 2018.
54. Zheng, X. L., J. L. Wang, L. F. Shangguan, Z. M. Zhou, and Y. H. Liu, "Smokey: Ubiquitous smoking detection with commercial Wi-Fi infrastructures," *Proc. 35th Annu. IEEE Int. Conf. Comput. Commun. (INFOCOM)*, 1–9, 2016.
55. Wu, C. S., Z. Yang, Z. M. Zhou, K. Qian, Y. H. Liu, and M. Y. Liu, "PhaseU: Real-time LOS identification with Wi-Fi," *Proc. IEEE INFOCOM*, 2038–2046, 2015.
56. Zhou, Z. M., Z. Yang, C. S. Wu, W. Sun, and Y. H. Liu, "LiFi: Line-of-sight identification with WiFi," *IEEE INFOCOM*, 2688–2696, 2014.
57. Wu, D., D. Q. Zhang, C. R. Xu, Y. S. Wang, and H. Wang, "WiDir: Walking direction estimation using wireless signals," *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput.*, 351–362, 2016.
58. Xi, W., J. Z. Zhao, X. Y. Li, K. Zhao, S. J. Tang, X. Liu, and Z. P. Jiang, "Electronic frog eye: Counting crowd using Wi-Fi," *IEEE INFOCOM 2014*, 361–369, 2014.
59. Depatla, S., A. Muralidharan, and Y. Mostofi, "Occupancy estimation using only Wi-Fi power measurements," *IEEE J. Sel. Areas Commun.*, Vol. 33, No. 7, 1381–1393, 2015.
60. Yang, Y. N., J. N. Cao, X. F. Liu, and X. L. Liu, "Wi-Count: Passing people counting with COTS Wi-Fi devices," *Proc. 27th Int. Conf. Comput. Commun. Netw. (ICCCN)*, 1–9, 2018.
61. Qian, K., C. S. Wu, Y. Zhang, G. D. Zhang, Z. Yang, and Y. H. Liu, "Widar2. 0: Passive human tracking with a single Wi-Fi link," *Proc. 16th Annu. Int. Conf. Mobile Syst. Appl. Services*, 350–361, 2018.
62. Poudel, K. N., D. Schurig, and N. Patwari, "Spatial imaging using a communication system's channel state information," *Proc. USNC-URSI Radio Sci. Meeting*, 41–42, 2016.
63. Vakalis, S., L. Gong, and N. A. Jeffrey, "Imaging with Wi-Fi," *IEEE Access*, Vol. 34, 28616–28624, 2019.
64. Wang, X. Y., C. Yang and S. W. Mao, "PhaseBeat: Exploiting CSI phase data for vital sign monitoring with commodity Wi-Fi devices," *Proc. IEEE 37th Int. Conf. Distrib. Comput. Syst. (ICDCS)*, 1230–1239, 2017.
65. Zeng, Y. W., D. Wu, J. Xiong, E. Yi, R. Y. Gao, and D. Q. Zhang, "FarSense: Pushing the range limit of Wi-Fi-based respiration sensing with CSI ratio of two antennas," *Proc. ACM Interact. Mobile Wearable Ubiquitous Technol.*, Vol. 3, No. 3, 1–26, 2019.
66. Wang, L., K. Sun, H. P. Dai, W. Wang, K. Huang, A. X. Liu, X. Y. Wang, and Q. Gu, "WiTrace: Centimeter-level passive gesture tracking using OFDM signals," *IEEE. Trans. Mob. Comput.*, Vol. 20, No. 4, 1730–1745, 2019.
67. Sun, L., S. Sen, D. Koutsonikolas, and K.-H. Kim, "Widraw: Enabling hands-free drawing in the air on commodity Wi-Fi devices," *Proc. 21st Annu. Int. Conf. Mobile Comput. Netw.*, 77–89, 2015.
68. Qian, K., C. S. Wu, Z. M. Zhou, Y. Zheng, Z. Yang, and Y. H. Liu, "Inferring motion direction using commodity wi-fi for interactive exergames," *Proc. CHI Conf. Human Factors Comput. Syst. (CHI)*, 1961–1972, 2017.
69. Wang, Y. X., K. S. Wu, and L. M. Ni, "Wifall: Device-free fall detection by wireless networks," *IEEE. Trans. Mob. Comput.*, Vol. 16, No. 2, 581–594, 2016.
70. Simone, D. D., D. S. Mauro, and C. Ernestina, and B. Giuseppe, "A trained-once crowd counting method using differential Wi-Fi channel state information," *Proc. 3rd Int. Workshop Phys. Analytics*, 37–42, 2016.
71. Liu, S. Q., Y. C. Zhao, and B. Chen, "WiCount: A deep learning approach for crowd counting using Wi-Fi signals," *Proc. IEEE Int. Symp. Parallel Distrib. Process. Appl. IEEE Int. Conf. Ubiquitous Comput. Commun. (ISPA/IUCC)*, 967–974, 2017.

72. Sharma, A., J. Y. Li, D. Mishra, G. Batista, and A. Seneviratne, "Passive Wi-Fi CSI sensing based machine learning framework for COVID-Safe occupancy monitoring," *IEEE ICC*, 1–6, 2021.
73. Gao, Q. H., J. Wang, X. R. Ma, X. Y. Feng, and H. Y. Wang, "CSI-based device-free wireless localization and activity recognition using radio image features," *IEEE Trans. Veh. Technol.*, Vol. 66, No. 11, 10346–10356, 2017.
74. Zhao, Z. Z., Z. Y. Lou, R. B. Wang, Q. Y. Li, and X. Xu, "I-WKNN: Fast-speed and high-accuracy WI-FI positioning for intelligent sports stadiums," *Comput. Electr. Eng.*, Vol. 98, 107619, 2022.
75. Ali, K., A. X. Liu, W. Wang, and M. Shahzad, "Recognizing keystrokes using Wi-Fi devices," *IEEE J. Sel. Areas Commun.*, Vol. 35, No. 5, 1175–1190, 2017.
76. Ali, K., A. X. Liu, W. Wang, and M. Shahzad, "Keystroke recognition using Wi-Fi signals," *Proc. 21st Annu. Int. Conf. Mobile Comput. Netw.*, 90–102, 2015.
77. Zhang, J., B. Wei, W. Hu, and S. S. Kanhere, "Wi-Fi-id: Human identification using Wi-Fi signal," *Proc. Int. Conf. Distrib. Comput. Sensor Syst.*, 75–82, 2016.
78. Xin, T., B. Guo, Z. Wang, M. Y. Li, Z. W. Yu, and X. S. Zhou, "Freesense: Indoor human identification with Wi-Fi signals," *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, 1–7, 2016.
79. Abdelnasser, H., M. Youssef, and K. A. Harras, "Wigest: A ubiquitous Wi-Fi-based gesture recognition system," *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, 1472–1480, 2015.
80. Melgarejo, P., X. Y. Zhang, P. Ramanathan, and D. Chu, "Leveraging directional antenna capabilities for fine-grained gesture recognition," *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput.*, 541–551, 2014.
81. He, W. F., K. S. Wu, Y. P. Zou, and Z. Ming, "WiG: WiFi-based gesture recognition system," *24th Int. Conf. on Computer Commun. and Networks (ICCCN)*, 1–7, 2015.
82. Wang, W., A. X. Liu, M. Shahzad, K. Ling, and S. L. Lu, "Device-free human activity recognition using commercial Wi-Fi devices," *IEEE J. Sel. Areas Commun.*, Vol. 35, No. 5, 1118–1131, 2017.
83. Arshad, S., C. H. Feng, Y. H. Liu, Y. P. Hu, R. Y. Yu, S. W. Zhou, and H. Li, "Wi-chase: A Wi-Fi based human activity recognition system for sensorless environments," *Proc. IEEE 18th Int. Symp. World Wireless Mobile Multimedia Netw. (WoWMoM)*, 1–6, 2017.
84. Wang, Y., J. Liu, Y. Y. Chen, M. Gruteser, J. Yang, and H. B. Liu, "E-eyes: Device-free location-oriented activity identification using fine-grained Wi-Fi signatures," *Proc. 20th Annu. Int. Conf. Mobile Comput. Netw.*, 617–628, 2014.
85. Chen, Z. H., L. Zhang, C. Y. Jiang, Z. G. Cao, and W. Cui, "Wi-Fi CSI based passive human activity recognition using attention based BLSTM," *IEEE. Trans. Mob. Comput.*, Vol. 18, No. 11, 2714–2724, 2018.
86. Yan, H., Y. Zhang, Y. J. Wang, and K. L. Xu, "WiAct: A passive Wi-Fi-based human activity recognition system," *IEEE Sens. J.*, Vol. 20, No. 1, 296–305, 2019.
87. Yousefi, S., H. Narui, S. Dayal, S. Ermon, and S. Valaee, "A survey on behavior recognition using Wi-Fi channel state information," *IEEE Commun. Mag.*, Vol. 55, No. 10, 98–104, 2017.
88. Li, H. J., X. He, X. K. Chen, Y. Y. Fang, and Q. Fang, "Wi-Motion: A robust human activity recognition using Wi-Fi signals," *IEEE Access*, Vol. 7, 153287–153299, 2019.
89. Xiao, F., J. Chen, X. H. Xie, L. Q. Gui, L. J. Sun, and R. C. Wang, "SEARE: A system for exercise activity recognition and quality evaluation based on green sensing," *IEEE Trans. Emerg. Top. Comput.*, Vol. 8, No. 3, 752–761, 2018.
90. Wang, W., A. X. Liu, M. Shahzad, K. Ling, and S. L. Lu, "Understanding and modeling of Wi-Fi signal based human activity recognition," *Proc. 21st Annu. Int. Conf. Mobile Comput. Netw.*, 65–76, 2015.
91. Hu, Y. Q., F. Zhang, C. S. Wu, B. B. Wang, and K. J. R. Liu, "DeFall: Environment-independent passive fall detection using Wi-Fi," *IEEE Internet Things J.*, Vol. 9, No. 11, 8515–8530, 2021.
92. Wang, H., D. Q. Zhang, Y. S. Wang, J. Y. Ma, Y. X. Wang, and S. J. Li, "RT-Fall: A real-time and contactless fall detection system with commodity Wi-Fi devices," *IEEE. Trans. Mob. Comput.*, Vol. 16, No. 2, 511–526, 2016.

93. Chen, C., Y. Han, Y. Chen, H. Q. Lai, F. Zhang, B. B. Wang, and K. J. R. Liu, "TR-BREATH: Time-reversal breathing rate estimation and detection," *IEEE Trans. Biomed. Eng.*, Vol. 65, No. 3, 489–501, 2017.
94. Guo, X. N., B. Liu, C. Shi, H. B. Liu, Y. Y. Chen, and M. C. Chuah, "Wi-Fi-enabled smart human dynamics monitoring," *Proc. 15th ACM Conf. Embedded Netw. Sensor Syst. (SenSys)*, 1–13, 2017.
95. Feng, C. H., S. Arshad, and Y. H. Liu, "Mais: Multiple activity identification system using channel state information of Wi-Fi signals," *Proc. 12th Int. Conf. Wireless Algorithms Syst. Appl. (WASA)*, 419–432, 2017.
96. Liu, X. F., J. N. Cao, S. J. Tang, J. Q. Wen, and P. Guo, "Contactless respiration monitoring via off-the-shelf Wi-Fi devices," *IEEE. Trans. Mob. Comput.*, Vol. 15, No. 10, 2466–2479, 2015.
97. Zeng, Y. Z. and P. H. Pathak and P. Mohapatra, "WiWho: Wi-Fi-based person identification in smart spaces," *Proc. 15th Int. Conf. Inf. Process. Sensor Netw.*, 1–12, 2016.
98. Wang, W., A. X. Liu, and M. Shahzad, "Gait recognition using Wi-Fi signals," *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput. (UbiComp)*, 363–373, 2016.
99. Gu, Y., X. Zhang, Z. Liu, and F. J. Ren, "BeSense: Leveraging Wi-Fi channel data and computational intelligence for behavior analysis," *IEEE Comput. Intell. Mag.*, Vol. 14, No. 4, 31–41, 2019.
100. Tan, S. and J. Yang, "Wi-Finger: Leveraging commodity Wi-Fi for fine-grained finger gesture recognition," *Proc. 17th ACM Int. Symp. Mobile Ad Hoc Netw. Comput.*, 201–210, 2016.
101. Li, H., W. Yang, J. X. Wang, Y. Xu, and L. S. Huang, "Wi-Finger: Talk to your smart devices with finger-grained gesture," *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput.*, 250–261, 2016.
102. Wang, G. H., Y. P. Zou, Z. M. Zhou, K. S. Wu, and L. M. Ni, "We can hear you with Wi-Fi!" *IEEE. Trans. Mob. Comput.*, Vol. 15, No. 11, 2907–2920, 2016.
103. Li, S. J., X. Li, Q. Lv, G. Y. Tian, and D. Q. Zhang, "Wi-Fit: Ubiquitous bodyweight exercise monitoring with commodity wi-fi devices," *Proc. IEEE SmartWorld, Ubiquitous Intell. Comput., Adv. Trusted Comput., Scalable Comput. Commun., Cloud Big Data Comput., Internet People Smart City Innov. (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*, 530–537, 2018.
104. Wright, J., A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 31, No. 2, 210–227, 2009.
105. Rao, B. D. and K. V. S. Hari, "Performance analysis of Root-Music," *IEEE Trans. Acoust., Speech, Signal Process.*, Vol. 37, No. 12, 1939–1949, 1989.
106. Hyvärinen, A. and E. Oja, "Independent component analysis: Algorithms and applications," *Neural Netw.*, Vol. 13, No. 4–5, 411–430, 2000.
107. Miyazaki, M., S. Ishida, A. Fukuda, T. Murakami, and S. Otsuki, "Initial attempt on outdoor human detection using IEEE 802.11ac WLAN signal," *Proc. IEEE Sensors Appl. Symp. (SAS)*, 1–6, 2019.
108. Takahashi, R., S. Ishida, A. Fukuda, T. Murakami, and S. Otsuki, "DNN-based outdoor NLOS human detection using IEEE 802.11ac WLAN signal," *Proc. IEEE Sensors*, 1–4, 2019.
109. Geng, C. H., X. Yuan, and H. Huang, "Exploiting channel correlations for NLOS ToA localization with multivariate Gaussian mixture models," *IEEE Wirel. Commun. Lett.*, Vol. 9, No. 1, 70–73, 2020.
110. He, J. H. M, S. Terashima, H. Yamada, and S. Kidera, "Diffraction signal-based human recognition in Non-Line-of-Sight (NLOS) situation for millimeter wave radar," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, Vol. 14, 4370–4380, 2021.
111. Wei, J. S., S. J. Wei, X. Y. Liu, M. Wang, J. Shi, and X. L. Zhang, "Non-Line-Of-Sight imaging by millimeter wave radar," *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, 2983–2986, 2021.
112. Huang, B. Q., G. Q. Mao, Y. Qin, and Y. Wei, "Pedestrian flow estimation through passive wifi sensing," *IEEE. Trans. Mob. Comput.*, Vol. 20, No. 4, 1529–1542, 2019.

113. Trivedi, A., C. Zakaria, R. Balan, A. Becker, G. Corey, and P. Shenoy, "Wifitrace: Network-based contact tracing for infectious diseases using passive wifi sensing," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, Vol. 5, No. 1, 1–26, 2021.
114. Mammen, P. M., C. Zakaria, T. Molom-Ochir, A. Trivedi, P. Shenoy, and R. Balan, "WiSleep: Scalable sleep monitoring and analytics using passive WiFi sensing," *arXiv preprint arXiv:2102.03690*, 2021.
115. Huang, Y. X., Q. Luo, S. Y. Ma, S. Hu, and Y. Gao, "Constant envelope OFDM RadCom system," *Proceedings of the International Conference on Communications, Signal Processing, and Systems*, 896–904, 2017.
116. Shi, C. G., F. Wang, M. Sellathurai, J. J. Zhou, and S. Salous, "Power minimization-based robust OFDM radar waveform design for radar and communication systems in coexistence," *IEEE Trans. Signal Process.*, Vol. 66, No. 5, 1316–1330, 2018.
117. Keskin, M. F., V. Koivunen, and H. Wymeersch, "Limited feedforward waveform design for OFDM dual-functional radar-communications," *IEEE Trans. Signal Process.*, Vol. 69, 2955–2970, 2021.
118. Kumbul, U., N. Petrov, F. van der Zwan, C. S. Vaucher, and A. Yarovoy, "Experimental investigation of phase coded FMCW for sensing and communications," *15 th European Conference on Antennas and Propagation (EuCAP)*, 1–5, 2021.
119. Uysal, F., "Phase-coded FMCW automotive radar: System design and interference mitigation," *IEEE Trans. Veh. Technol.*, Vol. 69, No. 1, 270–281, 2020.
120. McCormick, P. M., C. Sahin, S. D. Blunt, and J. G. Metcalf, "FMCW implementation of Phase-Attached Radar-Communications (PARC)," *IEEE Radar Conference (RadarConf)*, 1–6, 2019.
121. Chen, K., H. X. Zhang, Z. Y. Xu, and S. L. Pan, "FMCW lidar with communication capability using phase-diversity coherent detection," *Proc. 24th Optoelectron. Commun. Conf. (OECC) Int. Conf. Photon. Switching Comput. (PSC)*, 1–3, 2019.
122. Barrenechea, P., F. Elferink, and J. Janssen, "FMCW radar with broadband communication capability," *Proc. Eur. Radar Conf.*, 130–133, 2007.
123. Hadani, R., S. Rakib, M. Tsatsanis, A. Monk, A. J. Goldsmith, A. F. Molisch, and R. Calderbank, "Orthogonal time frequency space modulation," *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, 1–6, 2017.
124. Wei, Z. Q., W. J. Yuan, S. Y. Li, J. H. Yuan, G. Bharatula, R. Hadani, and L. Hanzo, "Orthogonal time-frequency space modulation: A promising next-generation waveform," *IEEE Wirel. Commun.*, Vol. 28, No. 4, 136–144, 2021.
125. Gaudio, L., M. Kobayashi, B. Bissinger, and G. Caire, "Performance analysis of joint radar and communication using OFDM and OTFS," *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, 1–6, 2019.
126. Wild, T., V. Braun, and H. Viswanathan, "Joint design of communication and sensing for beyond 5G and 6G systems," *IEEE Access*, Vol. 9, 30845–30857, 2021.
127. Cui, Y. H., X. J. Jing, and J. S. Mu, "Integrated sensing and communications via 5G NR waveform: Performance analysis," *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 8747–8751, 2022.
128. Liu, F., Y. H. Cui, C. Masouros, J. Xu, T. X. Han, Y. C. Eldar, and S. Buzzi, "Integrated sensing and communications: Towards dual-functional wireless networks for 6G and beyond," *IEEE J. Sel. Areas Commun.*, Vol. 40, No. 6, 1728–1767, 2022.
129. Liu, F., W. J. Yuan, C. Masouros, and J. H. Yuan, "Radar-assisted predictive beamforming for vehicular links: Communication served by sensing," *IEEE Trans. Wirel. Commun.*, Vol. 19, No. 11, 7704–7719, 2020.
130. Yuan, W. J., F. Liu, C. Masouros, J. H. Yuan, D. W. K. Ng, and N. González-Prelcic, "Bayesian predictive beamforming for vehicular networks: A low-overhead joint radar-communication approach," *IEEE Trans. Wirel. Commun.*, Vol. 20, No. 3, 1442–1456, 2021.
131. Liu, C., W. J. Yuan, S. Y. Li, X. M. Liu, H. S. Li, D. W. K. Ng, and Y. H. Li, "Learning-based predictive beamforming for integrated sensing and communication in vehicular networks," *IEEE J. Sel. Areas Commun.*, Vol. 40, No. 8, 2317–2334, 2022.

132. Mu, J. S., Y. Gong, F. P. Zhang, Y. H. Cui, F. Zheng, and X. J. Jing, "Integrated sensing and communication-enabled predictive beamforming with deep learning in vehicular networks," *IEEE Commun. Lett.*, Vol. 25, No. 10, 3301–3304, 2021.
133. Wang, X. Y, Z. S. Fei, J. A. Zhang, and J. X. Huang, "Sensing-assisted secure uplink communications with full-duplex base station," *IEEE Commun. Lett.*, Vol. 26, No. 2, 249–253, 2022.
134. Mei, W. D. and R. Zhang, "UAV-sensing-assisted cellular interference coordination: A cognitive radio approach," *IEEE Wirel. Commun. Lett.*, Vol. 9, No. 6, 799–803, 2020.